UNIVERSITÀ DELLA SVIZZERA ITALIANA
SWISS FINANCE INSTITUTE

DOCTORAL DISSERTATION

Essays in Asset Pricing

Author:
Piotr ORŁOWSKI

Supervisor:
Prof. Fabio TROJANI

A dissertation submitted in fulfillment of the requirements
for the degree of
PhD in Economics
SFI PhD in Finance

in the

Institute of Finance
Faculty of Economics

Dissertation Committee:

Prof. Patrick GAGLIARDINI, Università della Svizzera italiana

Prof. Olivier SCAILLET, Université de Genève

Prof. Paul SCHNEIDER, Università della Svizzera italiana

Prof. Fabio TROJANI, Université de Genève

July 10, 2017

Abstract

My dissertation consists of three chapters, each of which focuses on a different area of research in asset pricing. The first chapter's focal point is the measurement of the premium for jump risks in index option markets. The second chapter is devoted to non-parametric measurement of pricing kernel dispersion. The third chapter contributes to the literature on latent state variable recovery in option pricing models.

In the first chapter, *Big Risk*, I show how to replicate a large family of high-frequency measures of realised return variation using dynamically rebalanced option portfolios. With this technology investors can generate optimal hedging payoffs for realised variance and several measures of realised jump variation in incomplete option markets. These trading strategies induce excess payoffs that are direct compensation for second- and higher order risk exposure in the market for (index) options. Sample averages of these excess payoffs are natural estimates of risk premia associated with second- and higher order risk exposures. In an application to the market for short-maturity European options on the S&P500 index, I obtain new important evidence about the pricing of variance and jump risk. I find that the variance risk premium is positive during daytime, when the hedging frequency is high enough, and negative during night-time. Similarly, for an investor taking long variance positions, daytime profits are grater in absolute value than night-time losses. Compensation for big risk is mostly available overnight. The premium for jump skewness risk is positive, while the premium for jump quarticity is negative (contrary to variance, also during the trading day). The risk premium for big risk is concentrated in states with large recent big risk realisations.

In the second chapter, *Arbitrage free dispersion*, co-authored with András Sali and Fabio Trojani, we develop a theory of arbitrage-free dispersion (AFD) which allows for direct insights into the dependence structure of the pricing kernel and storck returns, and which characterizes the testable restrictions of asset pricing models. Arbitrage-free dispersion arises as a consequence of Jensen's inequality and the convexity of the cumulant generating function of the pricing kernel and returns. It implies a wide family of model-free dispersion constraints, which extend the existing literature on dispersion and co-dispersion bounds. The new techniques are applicable within a unifying approach in multivariate and multiperiod settings. In an empirical application, we find that the dispersion of stationary and martingale pricing kernel components in a benchmark long-run risk model yields a counterfactual dependence of short- vs. long-maturity bond returns and is insufficient for pricing optimal portfolios of market equity and short-term bonds.

In the third chapter, *State recovery from option data through variation swap rates in the presence of unspanned skewness*, I show that a certain class of variance and skew swaps can be thought of as sufficient statistics of the implied volatility surface in

the context of uncovering the conditional dynamics of second and third moments of index returns. I interpret the slope of the Cumulant Generating Function of index returns in the context of tradable swap contracts, which nest the standard variance swap, and share its fundamental linear pricing property in the class of Affine Jump Diffusion models. Equipped with variance- and skew-pricing contracts, I investigate the performance of a range of state variable filtering setups in the context of the stylized facts uncovered by the recent empirical option pricing literature, which underlines the importance of decoupling the drivers of stochastic volatility from those of stochastic (jump) skewness. The linear pricing structure of the contracts allows for an exact evaluation of the impact of state variables on the observed prices. This simple pricing structure allows me to design improved low-dimensional state-space filtering setups for estimating AJD models. In a simulated setting, I show that in the presence of unspanned skewness, a simple filtering setup which includes only prices of skew and variance swaps offers significant improvements over a high- dimensional filter which treats all observed option prices as observable inputs.

Acknowledgements

I can recall equally vividly the first day of the PhD program and the day of the dissertation defence, and thus the similar feelings of anxiety and hope at the beginning of this journey, and at its end, when another one begins. In the very limited space of this page, I want to thank the exceptional people whom I met on my doctoral student's journey. They influenced me professionally, academically and became a part of my circle of cherished friends.

I would not have made it without my dissertation advisor, Fabio Trojani. Fabio offered enthusiastic support for my scientific inquests, and patience when the stream of results was trickling slowly. I could always lean on his fundamental understanding of Finance and Econometrics which allowed me to take the right direction whenever I got too tangled in technical details to see the bigger picture. Thank you, Fabio!

I am indebted to Viktor Todorov for making my year-long visit at the Kellogg School of Management possible. The discussions with Viktor and seminars at Kellogg were very inspiring, and life near Chicago was a wonderful personal experience. I am again indebted to Viktor for supporting my academic job market efforts and believing in me as a researcher.

Learning how to teach might be the hardest part of a doctoral student's effort. I learned the most from Patrick Gagliardini, a great econometrician and teacher. He knows how to break up a complex problem into pills that a student can swallow and learn without realizing how complex the problem was in the first place.

I would like to thank the remaining members of my dissertation committee, Paul Schneider and Olivier Scaillet for discussions, support and encouragement in face of adversity.

Living and studying in Lugano would not have been the same without my colleagues and friends. András, Amina, Ilaria, Wojtek, Giuseppe, Stefania, Edoardo, Tamara, Alexandru, Elisa, Peter, thanks for discussions, technical assistance, hikes, board game nights, skiing trips and cycling from Lugano to San Bernardino pass and back. I hope that our paths will keep crossing in the future.

This dissertation is dedicated to Milena, my wonderful wife. I only met her because my doctoral studies took longer than they should have. I could not be happier. It is also dedicated to my parents, Bożena and Zygmunt, and to my sister, Beata. I can count on them in moments of despair and in moments of joy.

Contents

List of Figures

List of Tables

1 Big risk

1.1 Introduction

The financial distinction between run-of-the-mill price variation and "disastrous events" is usually made by introducing risks that scale with time (small risks) and risks that are significant even at minuscule decision horizons (big risks). Aït-Sahalia, 2004 succinctly states that "...the ability to disentangle jumps from volatility is the essence of risk management, which should focus on controlling large risks leaving aside the day-to-day Brownian fluctuations". In this context, jumps are naturally identified as big risk, and diffusive variation as small risk. While it is widely accepted in the literature that investors exhibit different attitudes to small and big risk, the possibility of disentangling them for risk sharing purposes has not yet been fully explored.

In this paper I introduce a general trading technology which allows for optimal hedging of a family of realised risk measures in incomplete option markets. The technology introduces two core innovations. First, it improves upon commonly employed techniques of hedging non-linear payoffs by providing a convenient optimal option portfolio for a given set of observed price quotes. Second, through a dynamic generalization, it renders many important measures of realised risk directly available to investors, as hedgeable payoffs.[1] Equipped with this technology, I analyse the traded properties of higher order risk in the market for one-week maturity European index options at the CBOE.[2] The CBOE is one of the world's most busy derivative trading venues, and the available data has enormous information content. I find that the compensation for variance risk is positive at such short horizons, and that in disaggregated results it is positive during market opening hours and negative during overnight periods, the daytime effect being larger in magnitude. Furthermore, I document the existence of compensation for directional jump risk, which is similarly concentrated in overnight periods, and increases substantially after the S&P500 index is subject to a large shock.

The market for very short maturity index options remains severely under-studied at a time when its importance is soaring. Despite the availability of data, most research to-date treated it with distrust and often discarded it. To a large extent that is because, option markets have been analysed through the lens of models, and the

[1] I give the exact sense in which the resulting trading strategies provide hedges in Section 1.2.

[2] At the time of writing, options with the weekly Friday settlement calendar constituted between 15% and 40% of trading volume in all SPX options. In 2016, the CBOE followed up on their success and introduced weekly options settled on Mondays and Wednesdays.

short end of the term structure turned out to be, to put it mildly, problematic to fit. Jump risk, however, has been considered necessary for explaining many key observations about the option market, and in the extant mathematical framework it is even more so at short maturities. The prices of weekly options, analysed with the use of my technology offer the sharpest insights into investors' perception of jump risk, unobscured by modelling assumptions. I exploit the availability of high-frequency records of trades and quotes and show how the intuition about realised variation measures can be translated to analysing option data.

Realised variation measures were quickly embraced by the asset pricing literature, which treated them as risk factors in attempts to resolve a number of asset pricing puzzles. Tradability of the newly introduced risk factors remained an open question. In cases like variance swaps, where a tradable representation is available, issues arise about the discrepancy between the trade's complete-market form, and its feasible incomplete-market implementation. This study is built on the premise of ensuring that the hedging strategies are the *exact* representations of the realised variation measures used for measuring risk premia for small and big risk in the sense that the payoffs of the strategies at settlement should be as close as possible to the realised variation calculated directly from high-frequency price records. Those settlement payoffs, taken together with the accumulated cost of establishing option positions, form excess payoffs.

I will illustrate the issues with an example based on the ubiquitous synthetic variance swap. In such a hedge, a single option position is established at its inception, and an associated trading strategy is implemented in the forward market for the underlying. At maturity, the accumulated forward and option positions are settled and the payoffs should in principle be equal to realised variance calculated directly from the forward prices. Once the cost of the initial option portfolio is subtracted, the analyst has an excess payoff on the table. I form the swaps with the approach prevailing in the literature and compare the resulting swap rates and payoffs with those yielded by my technology.[3] The left panel of Figure 1.1 shows the comparison of hedging errors, measured as the absolute value of the difference between realised variance and the traded payoff. In annualised terms, the often used approach yields absolute hedging errors ranging up to 1 percentage point in variance terms, while my technology reduces them by an order of magnitude. The right panel of Figure 1.1 plots the difference between the commonly calculated variance swap rate, and my approach. The commonly calculated swap rates are lower, on average. As a result, an unconditional estimate of the variance risk premium is 10% off when the standard technology is used. This is not innocuous: the unconditional estimate of the premium is biased. The main conclusion is that in the common approach the trading strategy *does not hedge realised variance*, but also a noise component which should be outside of the researcher's focus. The noise is uncorrelated with realised variance

[3] An exact description of the procedures is available in Section 1.4.

and the market return, but its magnitude increases with realised variance. Ensuring that the settlement payoff is as close as possible to the hedged realised variation measure gives the researcher confidence that she is indeed measuring a *meaningful* risk premium. In this paper I show that this question becomes crucial when measuring the premia for higher order risks, where issues related to replication accuracy are a first-order effect.

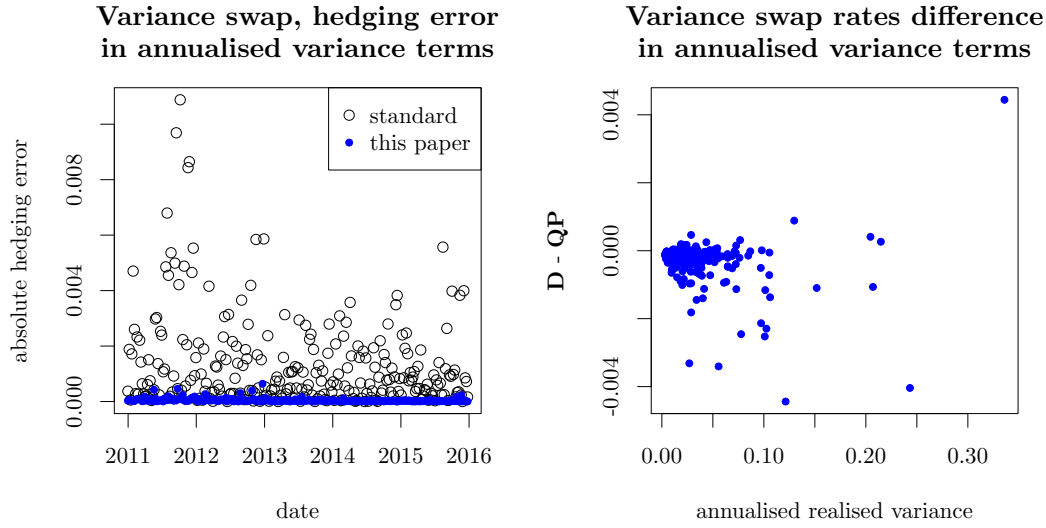


FIGURE 1.1: Trading technology: consequences of inadequate option hedging.

Jump variation proves especially significant when explaining the features of both stock and derivative markets, and many recent studies suggest that a premium for jump variation is the dominant component of both the equity and variance risk premia. In the literature, the standard approach to isolating a jump premium is via holding a delta-hedged, short-maturity, deep out-of-the-money option. The intrinsic difficulty of estimating such a risk premium lies in the fact that such strategies rarely pay off, yielding extremely skewed return distributions. The single option has to be sufficiently far out of the money as to avoid payoffs in cases of increased market volatility. Furthermore, this strategy does not hedge the investor against jump movements, that are subsequently reversed, so that ultimately the option expires worthless. To the contrary, the technology introduced in this paper allows the investor to hedge against adverse events on a continuous basis, through sequential rebalancing of the initial swap portfolio. In an event such as the Flash Crash on May 6^{th} 2010, the holder of an out-of-the-money, delta-hedged put would incur losses stemming from the delta-hedging error, and her hedge of the final asset value would expire worthless. An investor following the hedging strategy described in this paper would accrue a payoff when the market was collapsing, and possibly when it was recovering, depending on her hedging frequency.

The hedging technology has one additional advantage over a standard approach, exemplified before with variance swaps. Allowing for sequential option trades makes

it possible to hedge realised variation measures over arbitrary periods before option maturity while maintaining the benefit of exact hedging. I exploit this feature of the technology in order to investigate the difference in risk pricing between daytime and overnight trading. Investors holding positions in realised variation measures overnight are exposed to variance and higher order risk differently. While in daytime trading they can dynamically hedge or close the position altogether, overnight they face far more important constraints to trading, and much less efficient price discovery, if the market is open at all. The patterns that I find are striking. Long variance positions[4] are profitable, on average, during daytime trading, but in night-time trades they entail losses. Evidence on aggregate variance risk premia for maturities longer than two weeks showed that the premia were uniformly negative, which gave the variance swap the interpretation of an insurance contract. Not only is the insurance feature present exclusively in night-time trading, but also the daytime profits from long positions are greater than potential overnight losses: the aggregate variance risk premium over weekly horizons is positive. Where premia for jumps – big risk – are considered, similar patterns arise. Excess payoffs from long position in jump skewness and short in jump quarticity are significantly higher in overnight trades than in daytime trades. Additionally, the excess payoffs from such trades greatly increase after large stock price movements, even if subsequent big risks do not materialise. The effect of large innovations to the price of the index is thus clearly seen in a persistently elevated level of premia for big risk hedges.

In this paper I only consider hedging realised variation measures which are calculated indiscriminately for positive and negative returns. The technology is flexible enough, however, to extend the study towards exactly hedging variation measures such as realised semi-variances or third power of only negative jumps, for example. Such quantities merit extra attention, as evidenced by many studies of "good" and "bad" volatility.

1.1.1 Literature review

This paper is a direct off-shoot fo the literature on model-free hedging strategies. Since Neuberger, 1994 posited the creation of a log-return derivative, the literature gathered steam, culminating in several important studies. Jiang and Tian, 2005a implemented Britten-Jones and Neuberger, 2000's option-based estimator of model-free implied volatility, i.e. essentially the VIX^2 index. The most comprehensive, model-independent study of variance risk premia came from Carr and Wu, 2009a, who established the negative sign of the index variance risk premium, however used the standard approach to forming option portfolios, and calculated the floating leg of the swaps directly from the price process, instead of from the instrument payoffs; furthermore they noted that their floating leg corresponds only approximately

[4] A long position in a realised variation measure is a hedging strategy which at the settlement of the option and forward contracts pays the investor the accrued realised measure, if the latter is positive, and is paid by the investor otherwise.

to the theoretical option payoff in the presence of jump risk. Martin, 2012 proposed a change in definition of the option portfolio, and of the floating leg, this time calculated from payoffs, in order to obtain a variance swap whose swap leg exactly priced the floating leg even in the presence of jumps. Further on, Bondarenko, 2014 combined the previous approaches to develop a fully tradable setting, which also paves the way towards fully hedgeable skewness measures. All these approaches are generalised by the concept of realised divergence, and associated higher-order measures, introduced by Schneider and Trojani, 2015a. This paper takes the latter as a foundation to build upon, taking the realised divergence measures, recalled in Section 1.3.1, as primitives. A separate branch of the literature focused on providing pricing bounds for the prices of certain realised variation hedges, for example Hobson and Klimmek, 2012.

The semi-martingale model of asset prices has been the fundamental tool for defining and identifying small and big risks, and – in various guises – it was central for establishing the importance of big risks. The observation that stock returns are leptokurtic (e.g. Fama, 1965, who reports earlier findings by Moore, 1962) encouraged the first forays into jump modelling (see e.g. Press, 1967), based on estimation of simple semi-martingale models on low frequency data. With time, the frequency of returns and complexity of models and methods increased (Aït-Sahalia, 2004; Bates, 2012), and more model-agnostic methods for intra-day data, based on estimating quadratic variation, were developed (Barndorff-Nielsen and Shephard, 2006; Huang and Tauchen, 2005; Andersen, Bollerslev, and Huang, 2011).[5] Option price data has been used rarely in non-parametric studies (Li, 2011; Bollerslev and Todorov, 2011b; Bollerslev and Todorov, 2011a) and often in model-based investigations (Eraker, Johannes, and Polson, 2003; Andersen, Fusari, and Todorov, 2015b; Andersen, Bondarenko, and Gonzalez-Perez, 2015). All of these studies concluded that big risk is a significant contributor to the variation of asset prices (between 5 and 20% in most studies), and that big risk goes a long way towards explaining the dynamics of option prices. Lately, evidence arose from the analysis of intra-day data at tick frequencies which indicates that those estimates were inflated by an order of magnitude because of model mis-specification or erroneous classification of market movements as jumps, as in Christensen, Oomen, and Podolskij, 2014 and Bajgrowicz, Scaillet, and Treccani, 2015. Nevertheless, Christensen, Oomen, and Podolskij write that such erroneous classification is the consequence of *bursts in volatility*, when volatility suddenly rises and falls so fast, that on a given time scale continuous, yet abrupt price movements are indistinguishable from jumps. Thus, even though the latest evidence jump variation markedly diminishes its significance, it also shows that the *perception* of what constitutes a jump movement depends on the investors' profile. For a dynamic investor with a trading horizon of minutes or hours bursts

[5] The high-frequency methods do not posit parametric models for the stochastic volatility process or the jump distribution, but are built upon the foundation of the semi-martingale model.

in the drift or volatility of the underlying diffusion still constitute big risks. Furthermore, Liu, Longstaff, and Pan and Christensen, Oomen, and Podolskij point out the relation of big risk realizations to drops in liquidity, further emphasizing the inability to hedge them at the time adverse events hit.

A broad, yet more recent body of work investigates the pricing of big risk, and its importance in portfolio formation. Liu, Longstaff, and Pan, 2003 underline the infinitesimal impact of big risk on investor wealth, and the potential importance (and difficulty) of hedging jump risk with options. In studies of the cross section of stock returns, Bollerslev, Li, and Todorov, 2016 concluded that the exposure to small risks is essentially not compensated, while the exposure to big risk – of jump returns and overnight returns – is. Evidence from the aforementioned option pricing literature also indicates that jump risk carries significant premia.

While those studies offer a way of measuring big risk and shed some light on how investors price it, they offer little in way of hedging it separately from small risk. The access to sharing big risk is restricted in financial markets. Investors can purchase individual out-of-the-money, short-maturity options, however those do not protect against the *immediate* consequences of jumps. They can purchase variance swaps, which offer protection from immediate events, but do not discriminate between small and big risk. Through the dynamic option trading strategies one can achieve both objectives, at the price of exposing oneself to cumulative weighted risk of the price of volatility.

The remainder of this paper is laid out as follows. Section 1.2 makes precise the idea of tradable realised variation measures. In Section 1.3 I describe a family of tradable realised variation measures. In Section 1.4 I discuss the choice of strategies, the methods for rendering the exercise feasible, and the properties of the data set, as well as replication accuracy. I describe the empirical results in Section 1.6, and summarize the conclusions in Section 3.5.

1.2 Realized variation in asset pricing

In this section I stress the importance of obtaining *tradable* counterparts of the realized variation measures in order to ensure that the resulting estimates of higher order risk premia are meaningful.

With log returns $r_{t_j} : \ln F_j/F_{j-1}$ realized variance is defined as,

$$RV_{t,t+1} := \sum_{t \le t_j \le t+1} \ln^2 r_{t_j} \to \int_t^{t+1} \sigma_s^2 ds + \sum_{t \le s \le t+1} J_s^2 =: QV_{t,t+1}. \tag{1.2.1}$$

RV was initially a concept confined to the domain of financial econometrics. Over the course of twenty years of development, it gained importance in empirical asset pricing. Many studies have observed that the *variance risk premium*,

$$VRP_t := \mathbb{E}_t^{\mathbb{P}}\left[QV_{t,t+1}\right] - \mathbb{E}_t^{\mathbb{Q}}\left[QV_{t,t+1}\right]$$

has important predictability properties for both future volatility and index stock returns, and that it is important for the pricing of the cross-section of stocks. The rise of VRP to its current prominence was only possible because researchers found a way of approximately hedging the $QV_{t,t+1}$ payoff with extant financial instruments. The price of realized variance, $\mathbb{E}_t^{\mathbb{Q}}[QV_{t,t+1}]$, can be approximately calculated as the price of an option portfolio with readily calculable weights. The final payoff combines the option settlement with this of a series of positions in the underlying market. The VRP is typically available to investors through *variance swaps*, defined as paying realized daily squared log returns in exchange for a fixed amount, paid by the buyer of the swap upon inception. This approach is not exact, as explained in Carr and Wu, 2009b, and as I show below in Example 1, in the sense that there exists a discrepancy between what the seller of the swap is contractually obliged to pay the buyer, and what she can hedge in the market.

Nevertheless, since the success of realized variance as an asset pricing concept, many researchers shifted their interest towards studying other realized variation measures. At the same time, after Carr and Madan, 2001 showed how to replicate non-linear payoffs with European options, another strand of literature sprang up, investigating the relation between the prices of various non-linear replicating portfolios and the cross-section of stock returns (ANG et al., 2006; Chang, Christoffersen, and Jacobs, 2009; Goyal and Saretto, 2009; Amaya et al., 2011, , among others). These two branches of the literature should be meeting in the middle: finding prices that correspond to selected realized variation measures allows to calculate excess payoffs and estimate risk premia. This is because in empirical asset pricing, tradable quantities are of special importance, from the point of view of the financial theory (they allow for a direct representation of risk factors in the market). The concept of *alpha* as extraordinary returns beyond the fundamental compensation for risk is firmly based upon the idea that all risk factors under consideration are actually tradable. Unfortunately, it is not straightforward to engineer a trading strategy that replicates an arbitrary realized variation measure.

Even in the case of variance swaps, the literature relied heavily on approximations. The simplest – and most widely known – example of a variation-replicating strategy is the VIX^2-based variance swap. Typically, researchers interested in the variance risk premium calculate the payoffs of such swaps as

$$RV_{t,T} - VIX_{t,T}^2. \tag{1.2.2}$$

In pure diffusion models the $VIX_{t,T}^2$ is the exact price of $RV_{t,T}$. However, in the presence of jumps there is a wedge between what the option portfolio prices, and the limit of the floating leg of the swap. The wedge can be eliminated by slightly tweaking the definition of the floating leg, or equivalently, of the hedged realised variation measure. In order to put the VIX^2-based variance swap firmer on asset pricing footing, I present Schneider and Trojani's example of how the floating leg, or

equivalently, realized measure, should be defined so that the variance swap is *exactly* priced by the VIX^2 option portfolio.

Example 1 (The exact VIX^2 variance swap)**.** Schneider and Trojani, 2015a show that an option portfolio with long positions $\phi(K)$ in *out of the money* options (with prices $O_T(K)$ at strike K):

$$\phi(K) = \frac{2}{K^2}, \text{ price } \int_0^\infty \phi(K) O_T(K) dK =: VIX^2 \text{ and payoff } -2\left(\ln F_T - \ln F_0 - \frac{1}{F_0}(F_T - F_0)\right), \tag{1.2.3}$$

coupled with the dynamic trading strategy:

$$\sum_{j=1}^{n} \underbrace{\left(\frac{2}{F_{j-1}} - \frac{2}{F_0}\right)}_{\text{trading weights}} \underbrace{(F_j - F_{j-1})}_{\text{forward position}}, \tag{1.2.4}$$

pays at maturity the difference between the Itakura and Saito, 1968 divergence of the stock price the price of the initial option portfolio:

$$\underbrace{\sum_{j=1}^{n} 2\left[-\ln F_j/F_{j-1} + (F_j/F_{j-1} - 1)\right]}_{\text{realized divergence}} - \int_0^\infty \phi(K) O_T(K) dK. \tag{1.2.5}$$

The continuous-time limit of the divergence inside the summation in (1.2.5) is

$$\int_0^T \sigma_s^2 ds + \sum_{0 \le s \le T} 2\left[-\ln F_s/F_{s-} + (F_s/F_{s-} - 1)\right],$$

and the jump terms are of leading order $\ln^2 F_s/F_{s-}$. The divergence in the summation (1.2.5) is the sum of the option payoff in (1.2.3) and accumulated forward payoffs (1.2.4). The strategy defined by (1.2.3)-(1.2.4) is *model independent*: the replication argument is valid for any no-arbitrage market set-up and the continuous-time limit is valid in a semi-martingale setting. Equation (1.2.3) is the definition of the square of the VIX index (CBOE, 2000) and as a consequence VIX-based variance swaps offer exact replication of realised divergence.

Example 1 is firmly grounded in a complete option market setting, i.e. it assumes that a continuum of options is purchased or sold at the inception of the hedging strategy. In reality, an investor at the CBOE has to form a replicating portfolio for the non-linear payoff with a finite (albeit large) number of options. In the introduction I already hinted at the fact that the choice of the incomplete market replicating option portfolio is not trivial, and may have significant consequences for the measured risk premium. I defer further consideration of this subject to Section 1.4.

1.2.1 Dynamic tradability

The trading strategy in Example 1 consists of a single option trade, at inception, accompanied by sequential trading in the (forward on the) underlying asset. Although this framework offers many interesting extensions, in this paper I advocate the use of strategies involving *both* dynamic option and forward trading. I start from two fundamental results. First, Carr and Madan's replication formula, which is essential to understand how the derivatives are traded when the investor wants to replicate a realized variation measure. Second, from the observation by Schneider and Trojani, 2015a that Carr and Madan's approach is strongly related to the information-theoretic concept of *divergence*.

At time t_j in a complete option market[6] with maturity T, for an arbitrary twice-differentiable function $g : \mathbb{R}_+ \to \mathbb{R}$, it is possible to form an option portfolio paying *at maturity* the Bregman, 1967 divergence of g:

$$\begin{aligned} G(F_T, F_{t_j}) :&= \int_0^\infty \left[g''(K)(K - F_T)\mathbf{1}_{\{F_T \leq F_{t_j}\}} + g''(K)(F_T - K)\mathbf{1}_{\{F_T \geq F_{t_j}\}} \right] dK \\ &= g(F_T) - g(F_{t_j}) - g'(F_{t_j})(F_T - F_{t_j}). \end{aligned} \tag{1.2.6}$$

The option portfolio is comprised of positions of size $g''(K)dK$ in put (call) options for $K \leq F_{t_j}$ ($K > F_{t_j}$). The price of the non-linear payoff is calculated simply as a weighted average of observed option prices:

$$\mathcal{G}_{t_j} := \int_0^\infty g''(K) O_{t_j}(K) dK. \tag{1.2.7}$$

A long position (i.e. a purchase of options with positive weights, and sale of options with negative weights) in the option portfolio pays exactly (1.2.6) at maturity.

Equipped with this concept of an option trade, I define the dynamic strategy. Let δ, ϕ be sequences of real numbers, potentially depending on F_{t_j} or other information available at time t_j.[7] With these tools one can form *dynamic option trading strategies*

$$\sum_{j=0}^{n-1} \phi_{t_j} \left[G(F_T, F_{t_j}) - \mathcal{G}_{t_j} \right] + \delta_{t_j}(F_T - F_i). \tag{1.2.8}$$

$\mathcal{G}_{t_j}$ was defined as the (t_j, T) forward price of $G(F_T, F_{t_j})$, and under no-arbitrage[8] there exists a forward-neutral measure $\mathbb{Q}_T$ such that $\mathcal{G}_{t_j} = \mathbb{E}^{\mathbb{Q}_T}\left[G(F_T, F_{t_j}) | \mathcal{F}_{t_j} \right]$.

[6] If a continuum of options on the forward is available with strikes ranging from 0 to ∞.

[7] Further in the text I consider continuous-time limits of the trading strategies. In order for them to exist, technical conditions have to be imposed on δ and ϕ, such as adaptedness and boundedness.

[8] Acciaio et al., 2013 show conditions for the existence of $\mathbb{Q}_T$ and absence of arbitrage.

I decompose the payoff of (1.2.8) into two components. The *settlement payoff*

$$\sum_{j=0}^{n-1} \phi_{t_j} G(F_T, F_{t_j}) + \delta_{t_j}(F_T - F_{t_j}), \tag{1.2.9}$$

consists of the trading gains in the forward market and the settlement payoffs of the option positions. The *aggregate cost*

$$\sum_{j=0}^{n-1} \phi_{t_j} \mathcal{G}_{t_j} \tag{1.2.10}$$

represents the total financial outlay, uncertain at time t_0, that the investor has to make in order to obtain the settlement payoff. The distinction is important: I design strategies whose settlement payoff corresponds exactly to realized variation measures in the following sense:

$$\sum_{j=0}^{n-1} \phi_{t_j} G(F_T, F_{t_j}) + \delta_{t_j}(F_T - F_{t_j}) = \sum_{j=0}^{n-1} G(F_{t_{j+1}}, F_{t_j}) \tag{1.2.11}$$

The VIX^2-based variance swap is a special case of strategy (1.2.8) with $\phi_{t_j} = 0$ for $j > 0$. The price for such a generalization of option trading is the blurring of the distinction between the fixed and floating legs of the variance swap contract. The strategy is, however, the *only* currently known way of obtaining equality in (1.2.11) with the use of two categories of financial instruments: European options and forwards. The concept of strategy costs returns into focus: it's the financial outlay necessary for obtaining the requisite settlement payoff.

By construction, the replication cost is not known at time t_0, but instead is stochastic. There are two sources of risk determining the aggregate option cost: the changes in weights ϕ_{t_j} and the changes in option prices. I give an interpretation of both in Section 1.3.2. Thus overall, dynamic option trading strategies can be understood as exchanges of the risk of the settlement payoff for the risk of the aggregate replication cost. In this setting I take the natural definition of the risk premium associated with replicating $\sum_{j=0}^{n-1} G(F_{t_{j+1}}, F_{t_j})$ at settlement to be:

$$\begin{aligned} &\mathbb{E}\left[\sum_{j=0}^{n-1} G(F_{t_{j+1}}, F_{t_j}) - \sum_{j=0}^{n-1} \phi_{t_j} \mathcal{G}_{t_j}\right] \\ &= \mathbb{E}\left[\sum_{j=0}^{n-1} \phi_{t_j} G(F_T, F_{t_j}) + \delta_{t_j}(F_T - F_{t_j}) - \sum_{j=0}^{n-1} \phi_{t_j} \mathbb{E}^{\mathbb{Q}_T}\left[G(F_T, F_{t_j}) | \mathcal{F}_{t_j}\right]\right], \end{aligned} \tag{1.2.12}$$

i.e. the final settlement payoff of the aggregate derivative position less the cost of "producing" the payoff. Such excess payoffs can indeed be useful for evaluating

asset pricing models. Write:

$$\mathbb{E}^{\mathbb{Q}_T}\left[\sum_{j=0}^{n-1}\phi_{t_j}\mathbb{E}^{\mathbb{Q}_T}\left[G(F_T,F_{t_j})|\mathcal{F}_{t_j}\right]\right]=\mathbb{E}\left[\sum_{j=0}^{n-1}\phi_{t_j}\mathbb{E}\left[\prod_{k=0}^{j-1}\frac{M_{t_{k+1}}}{M_{t_k}}G(F_T,F_{t_j})|\mathcal{F}_{t_j}\right]\right], \tag{1.2.13}$$

and note that in order to evaluate this expression non-parametrically it would suffice that at time t_0 options for each maturity t_j were quoted. Even in absence of such an abundance of data, (1.2.12) or (1.2.13) can be of use for asset pricers willing to characterise the higher moments of returns on assets such as hedge fund portfolios. Both the settlement payoff and the aggregate cost components of such strategies reflect all the risk factors to which the investors are exposed, contrary to resorting to using a realised variation measure as the driving force of the stochastic discount factor.

1.3 Tradable realized variation measures

In the previous Section I introduced the general concept of dynamic option trading, with the objective of replicating certain realised variation measures. That framework can be generalized even further, however I defer this discussion to the end of this Section. The level of generality is already sufficient to consider trading big risk separately from small risk. To fix ideas, before I move to defining the exactly replicable realized variation measures, I start with considering those types of realized variation measures that in a semi-martingale model separate jumps from Brownian increments, and those that do not.

If the asset price indeed follows a semi-martingale with finite-activity jumps, it is relatively easy to estimate hypothetical jump and Brownian variation separately. Realized variance, as defined in (1.2.1), is an example of a measure that aggregates both kind of risk. On the pure-jump side, the literature offers both indirect measures, such as a difference between realised variance and multi-power variation (Barndorff-Nielsen, Shephard, and Winkel, 2006), truncated measures (Mancini, 2001) or direct measures, such as 3rd or 4th power variation of log returns (Jacod and Protter, 2012):

$$RV_{t,t+1}:=\sum_{t\le t_j\le t+1}\ln^2 r_{t_j}\longrightarrow\int_t^{t+1}\sigma_s^2ds+\sum_{t\le s\le t+1}J_s^2 \tag{1.3.1}$$

$$RV_{t,t+1}-BV_{t,t+1}:=\sum_{t\le t_j\le t+1}\ln^2 r_{t_j}-\sum_{t\le t_j\le t+1}\left|\ln r_{t_j}\right|\left|\ln r_{t_{j-1}}\right|\longrightarrow\sum_{t\le s\le t+1}J_s^2 \tag{1.3.2}$$

$$RPV^3:=\sum_{j=1}^{n}\ln^3 r_{t_j}\longrightarrow\sum_{t_0\le s<t_n}J_s^3 \tag{1.3.3}$$

$$RPV^4:=\sum_{j=1}^{n}\ln^4 r_{t_j}\longrightarrow\sum_{t_0\le s<t_n}J_s^4. \tag{1.3.4}$$

These realized measures are not directly replicable, i.e. there does not exist a strategy of type (1.2.8), whose settlement payoff is equal to any of the quantities on the left-hand side of the limits in equations (1.3.1) through (1.3.4).

This Section is devoted to describing the family of realized variation measures which is replicable as settlement payoffs of dynamic option trading strategies introduced in the previous Section, and to a closer look at the behaviour of the aggregate option cost of the most important of such strategies.

1.3.1 Realised (jump) divergence

The starting point is the construction of realised weighted power divergence, a concept which generalises realised variance. Realised power divergence and associated power divergence swaps were introduced and comprehensively described in Schneider and Trojani, 2015a. Hereby I recall the basic definitions and properties.

Definition 1 (Realised power divergence). *Let F_t denote the forward price of the underlying asset at time t and maturing at time T and let γ_t define an adapted process. Set a grid of $n+1$ times, $0 = t_0 \leq t_1 \leq \ldots \leq t_n \leq T$. For function $\phi_p(x) := \frac{x^p-1}{p(p-1)}$ and $p \in \mathbb{R}$,[9] define realised power divergence as:*

$$D_\gamma^{n,p}(F) := \sum_{j=1}^{n} \gamma_{t_{j-1}} D_p(F_{t_j}, F_{t_{j-1}}) := \sum_{j=1}^{n} \gamma_{t_{j-1}} \frac{F_{t_j}^p - F_{t_{j-1}}^p}{p(p-1)} - \gamma_{t_{j-1}} \frac{F_{t_{j-1}}^{p-1}}{p-1} \left(F_{t_j} - F_{t_{j-1}}\right).$$

The weighting process γ plays an important role. Throughout this paper I concentrate on the fundamental case $\gamma_t := F_t^{-p}$, so that

$$\gamma_s D_p(F_t, F_s) = \frac{e^{p\ln(F_t/F_s)} - 1}{p(p-1)} - \frac{e^{\ln(F_t/F_s)} - 1}{p-1} = \frac{\ln^2(F_t/F_s)}{2} + O\left(\ln^3(F_t/F_s)\right). \tag{1.3.5}$$

Power divergence, for the appropriate γ scaling process, is locally quadratic in log returns, and insensitive to the *level* of F. Furthermore, the leading order of the Taylor expansion in (1.3.5) does not depend on p. Taking p-derivatives of (1.3.5) eliminates the leading order terms and yields higher-order realised measures.

Definition 2 (Realised jump divergence). *Let $\gamma^J = \gamma_t F_t^{-p}$ and γ_t does not depend on F nor p. For $p \in \mathbb{R}$ define realised jump skewness*

$$S_{\gamma^J}^{n,p}(F) := \frac{\partial D_{\gamma^J}^{n,p}(F)}{\partial p} = \sum_{j=1}^{n} \gamma_{t_{j-1}} \frac{\partial}{\partial p} \frac{D_p(F_{t_j}, F_{t_{j-1}})}{F_{t_{j-1}}^p} = \sum_{j=1}^{n} \gamma_{t_j} \frac{\ln^3(F_t/F_s)}{6} + O\left(\ln^4(F_t/F_s)\right),$$

[9] For $p = 0$ and $p = 1$ limits of all involved expressions exist.

and realised jump quarticity:[10]

$$Q^{n,p}_{\gamma^J}(F) := \frac{\partial^2 D^{n,p}_{\gamma^J}(F)}{\partial p^2} = \sum_{j=1}^{n} \gamma_{t_{j-1}} \frac{\partial^2}{\partial p^2} \frac{D_p(F_{t_j}, F_{t_{j-1}})}{F^p_{t_{j-1}}} = \sum_{j=1}^{n} \gamma_{t_j} \frac{\ln^4(F_t/F_s)}{12} + O\left(\ln^5(F_t/F_s)\right)$$

Khajavi, Orłowski, and Trojani, 2016 provide the theoretical framework for inference about realised (jump) divergence in a semi-martingale framework. Realised jump divergence measures in Definition 2 have well-defined high-frequency limits and their estimators based on discretized observations obey central limit theorems. Finally, I fix the idea of what the former realised variation measures represent in a semi-martingale framework.

Interpretation (1) (Continuous time limits)**.** Let F follow a general semi-martingale process with finite activity jumps:

$$\frac{dF_s}{F_{s-}} = \mu_s ds + \sigma_s dW_s + \int_{\mathbb{R}\setminus\{0\}} (e^x - 1)\, \nu_s(dx, dt). \tag{1.3.6}$$

Let $\mathbb{F}$ be the filtration to which the forward price process F is adapted, potentially greater than the one generated by F itself. Let γ be adapted to $\mathbb{F}$ and bounded. Let the maturity T be fixed and let $n \to \infty$ such that $\max_{i\in\{0\ldots n-1\}} |t_{i+1} - t_i| \to 0$. The following limit holds under technical assumptions in Khajavi, Orłowski, and Trojani, 2016:

$$\lim_{n\to\infty} D^{n,p}_{\gamma}(F) = \frac{1}{2}\int_0^T \gamma_s F^p_s \sigma^2_s ds + \sum_{0\le s\le T} \gamma_{s-} F^p_{s-} D_p(F_s, F_{s-}) \tag{1.3.7}$$

$$\lim_{n\to\infty} D^{n,p}_{F^{-p}_s} = \frac{1}{2}\int_0^T \sigma^2_s ds + \sum_{0\le s\le T} \frac{D_p(F_s, F_{s-})}{F^p_{s-}}. \tag{1.3.8}$$

The measures $S^{n,p}_{\gamma^J}(F)$ and $Q^{n,p}_{\gamma^J}(F)$ have the following pure jump limits:

$$\lim_{n\to\infty} S^{n,p}_{\gamma F^{-p}_s}(F) = \sum_{t_0\le s<T} \gamma_{s-} S_{F^{-p}_{s-}}(F_s, F_{s-}) \tag{1.3.9}$$

$$\lim_{n\to\infty} Q^{n,p}_{\gamma F^{-p}_s}(F) = \sum_{t_0\le s<T} \gamma_{s-} Q_{F^{-p}_{s-}}(F_s, F_{s-}) \tag{1.3.10}$$

In Figure 1.2 I illustrate how the divergence, skewness and quarticity functions behave for $p \in \{0, 1/2, 1\}$, when setting $F_0 = 1$ and varying F_T (the first argument). The bottom row of the Figure replicates the top row for a narrower range of F values. Power divergences clearly exhibit locally quadratic behaviour, skewness – locally cubic, while quarticity – locally quartic. The differences between payoff function values become apparent mostly for large deviations of F_T from 1, that is for large returns. In Figure 1.3 I demonstrate how realised skewness $S^{n,1/2}_{F_t^{-1/2}}(F)$ picks up

[10] Full formulae are given in equations (A.1.1) through (A.1.6) in Appendix A.1.1.A.1.1.

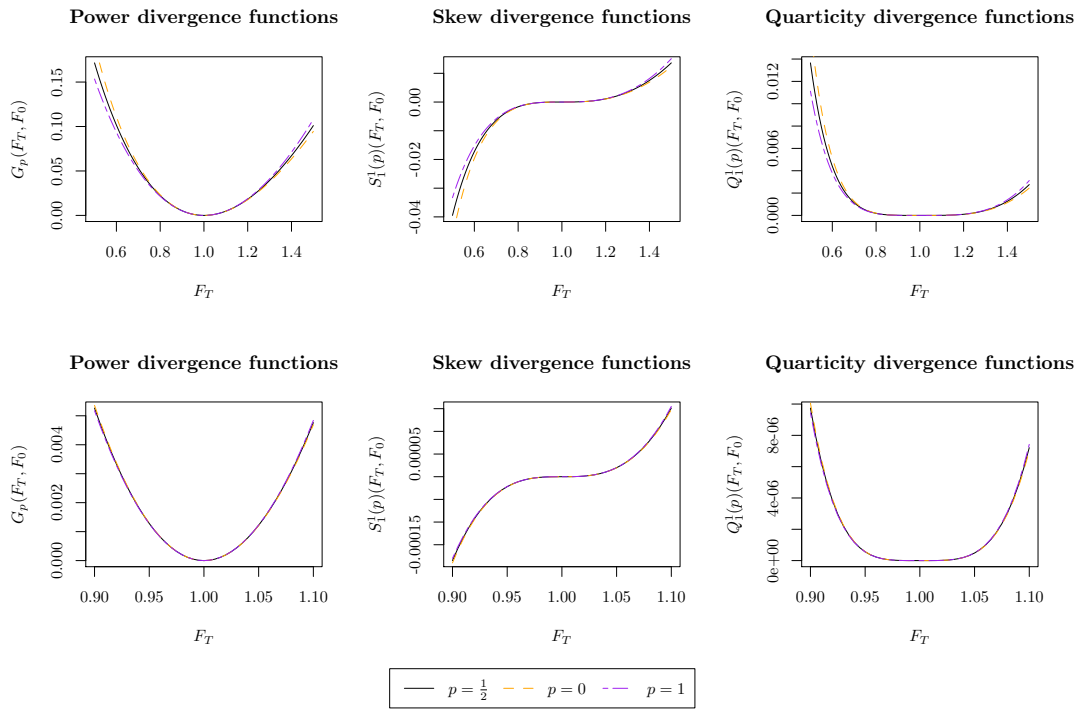


FIGURE 1.2: Payoffs: power divergence, skewness divergence, and quarticity divergence. Power divergence G_p plotted for $p \in \{1/2, 0, 1\}$. Skew and quarticity divergences formed around the respective power divergence function.

jumps in a simulated data set. The leftmost panel presents a high-frequency simulated price path. The top right panel presents log returns calculated from the path and three return jumps (marked with dashed orange lines) are clearly seen in the sample. The bottom right panel shows the cumulative skewness divergence during the trading period: the measure markedly decreases at each jump movement.

It is possible to design trading strategies that replicate realised jump divergence measures for arbitrary γ weighting strategies thanks to three fundamental results. First, Schneider and Trojani, 2015a show that the difference $D_p(F_T, F_s) - D_p(F_T, F_t)$ is an affine function of $F_T - F_t$, for $t > s$. Second, by linearity of differentiation, so are the p-derivatives of the difference, which define realised jump divergence measures. Third, Carr and Madan, 2001 show how to construct option portfolios replicating non-linear payoffs such as $D_p(F_T, F_s)$, $S_p(F_T, F_s)$ and $Q_p(F_T, F_s)$.

The exact replication of realised divergence measures associated with function g and its Bregman divergence G is possible if, as noted before, the difference between divergences is affine in $F_T - F_t$. Schneider and Trojani, 2015a showed that this is indeed the case for power divergences generated by function $\phi_p(x)$, $D_\gamma^{n,p}(F)$. For

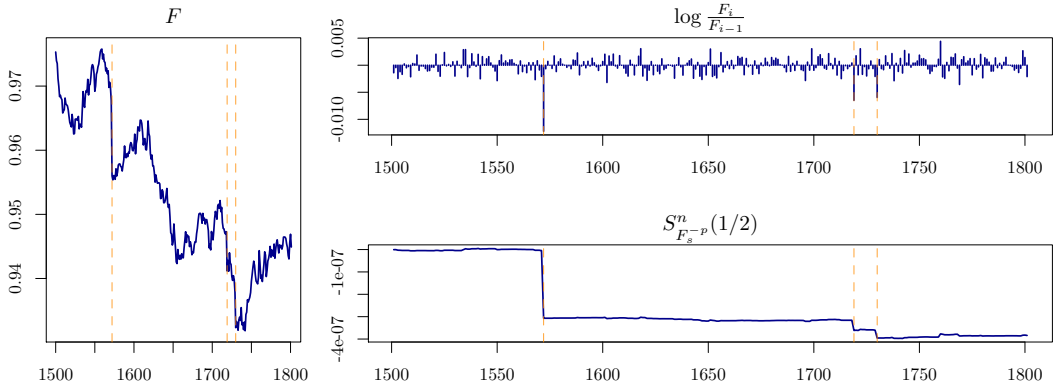


FIGURE 1.3: $S^n_{F_t^{-p}}(1/2)$: simulated data. The figure plots data from a simulated trading day at a 5-minute frequency. Three jumps (marked by vertical dashed orange lines) occur within the day. The left panel presents the evolution of the underlying price F. The top right panel plots the logarithmic returns. The bottom right panel plots cumulative Hellinger skewness calculated over the course of the trading day.

any realised divergence G meeting that condition, one can rewrite it – or the corresponding settlement payoff – as:

$$
\begin{aligned}
G^n_\gamma(F) &= \sum_{j=1}^{n} \gamma_{t_{j-1}} G\left(F_{t_j}, F_{t_{j-1}}\right) \qquad (1.3.11)\\
&= \sum_{j=1}^{n} \gamma_{t_{j-1}} \left[G\left(F_T, F_{t_{j-1}}\right) - G\left(F_T, F_{t_j}\right)\right] + \sum_{j=1}^{n} \gamma_{t_{j-1}} \left[g'_p(F_{t_{j-1}}) - g'_p(F_{t_j})\right]\left(F_T - F_{t_j}\right)\\
&= \gamma_{t_0} G\left(F_T, F_{t_0}\right) + \sum_{j=1}^{n} \left[\gamma_{t_j} - \gamma_{t_{j-1}}\right] G(F_T, F_{t_j}) - \gamma_{t_n} G\left(F_T, F_{t_n}\right)\\
&+ \sum_{j=1}^{n} \gamma_{t_{j-1}} \left[g'_p(F_{t_{j-1}}) - g'_p(F_{t_j})\right]\left(F_T - F_{t_j}\right).
\end{aligned}
$$

The dynamic trading strategy (1.2.8) whose *settlement payoff* is equal to (1.3.11) is sonsequently defined by ϕ such that $\phi_{t_0} = \gamma_{t_0}$, $\phi_{t_j} = \gamma_{t_j} - \gamma_{t_{j-1}}$ and $\phi_{t_n} = -\gamma_{t_n}$, and δ such that $\delta_{t_0} = 0$ and $\delta_{t_j} = \gamma_{t_{j-1}}\left[g'_p(F_{t_{j-1}}) - g'_p(F_{t_j})\right]$. The strategy can be implemented at an arbitrary frequency and its *exactness* for all considered realised measures is purely a consequence of the fact that equation (1.3.11) is an algebraic transformation of a high-frequency realised variation measure into expressions which can be obtained as settlement payoffs of option and forward positions assumed over time. A straightforward application of this basic strategy allows to trade weigthted realised power divergence. An extension to jump divergences requires some tedious algebra. An inspection of formulae (A.1.1) through (A.1.6) in Appendix A.1.1.A.1.1 indicates that the requisite strategies are more complex in two dimensions: first, these realised measures are combinations of four components: $F_T/F_s - 1$ (a forward position), $F_s^{-p} D_p(F_T, F_s)$ (power divergence weighted by F_s^{-p}), $K^{(1)}_{s,p} \ln(F_T/F_s)(F_T/F_s)^p$ and $K^{(2)}_{s,p} \ln^2(F_T/F_s)(F_T/F_s)^p$; second, equation (1.3.11) is not directly applicable to the two latter components. Some manipulation is required

to express them as *portfolios* of divergences with additional weighting. I defer full expressions to Appendix A.1.1.A.1.1, equations (A.1.7) through (A.1.12).

The variation measures $D^{n,p}_{\gamma^J}(F)$, $S^{n,p}_{\gamma^J}(F)$ and $Q^{n,p}_{\gamma^J}(F)$ are replicable as settlement payoffs (1.2.9) of portfolios of dynamic option trading strategies (1.2.8). The strategy position processes (ϕ, δ) can be calculated by applying (1.3.11) to equations (A.1.7) through (A.1.12). An investor who enters such a strategy faces outlays or proceeds from dynamic option trading (1.2.10) which are uncertain at time t_0. The investor can decide to replicate the payoff until time $t_n < T$, i.e. she can finish accruing the settlement payoff before the option maturity. In this case, from time t_n to time T she *holds an option position* and a *forward position* such that their payoffs at settlement will exactly offset each other, and while the payoff of any single instrument in the portfolio is not known before T, the aggregate payoff *is known*.

The theory of realized divergence provides realized variation measures whose interpretation is – to first order – identical will familiar power variation measures. The theory exploits the small-time difference between small and big risk to develop measures which isolate big risk. The great benefit of these measures is that – as shown above – there exist dynamic trading strategies (1.2.8), which yield a settlement payoff equivalent to the realised measure in the sense defined in equation (1.2.11).

1.3.2 Costs of dynamic replication

In the dynamic strategy (1.2.8) the uncertain option cost component (1.2.10) merits a closer look. The uncertainty means that an investor willing to obtain at settlement one of the payoffs under consideration is exposed to risks of changing position weights and changing option prices. In this section I investigate the most salient features of the aggregate cost risks for weights $\gamma = F_t^{-p}$ (scale-free divergence). The most important findings are that *a*) the option rebalancing costs for $S^{n,p}_{\gamma^J}(F)$ and $Q^{n,p}_{\gamma^J}(F)$ are not 0 even if the realised measures are almost surely equal to 0, i.e. no jumps are allowed in the model, *b*) the option rebalancing costs for $S^{n,p}_{\gamma^J}(F)$ are not 0 even if the price of a skewness swap is 0, and *c*) the rebalancing costs in each period are of the order $\ln(F_{t_i}/F_{t_{i-1}})\mathcal{D}_{p,t_j}$, i.e. they depend on return realisations, the price of divergence swaps, and the remaining maturity of the strategy.

The aggregate cost of implementing the dynamic trading strategy (1.2.8) is easiest to infer from the second line of (1.3.11):

$$\begin{aligned} \mathcal{C}[G^n_\gamma] &:= \sum_{j=1}^{n} \gamma_{j-1} \left[\mathbb{E}^{\mathbb{Q}_T}_{t_{j-1}} \left[G(F_T, F_{t_{j-1}}) \right] - \mathbb{E}^{\mathbb{Q}_T}_{t_j} \left[G(F_T, F_{t_j}) \right] \right] \\ &\xrightarrow{\mathbb{P}} \int_{t_0}^{t_n} -\gamma_s d\mathcal{G}_s, \end{aligned} \tag{1.3.12}$$

with $\mathcal{G}_t$ defined in equation (1.2.7). The option trading costs (or proceeds) are an integral of the weighting function with respect to the changes in the price of the

option portfolio that replicates the payoff $G(F_T, F_t)$. The option portfolio price $\mathcal{G}_t$ changes, firstly, because F_t changes (G itself is not necessarily scale-free), secondly, because of changes in relevant state variables that drive the forward price F (1.3.6) and directly influence option prices (e.g. stochastic volatility), and thirdly, because of the shrinking maturity of the traded contracts.

In this section I analyse the trading costs of a long strategy[11] in the payoff $D_\gamma^{n,p}$ with weighting process $\gamma = F_t^{-p}$ (see Definition 1), and I defer $S_{\gamma^J}^{n,p}$ and $Q_{\gamma^J}^{n,p}$ to the Appendix. The time-t swap rate for power divergence $D_p(F_T, F_t)$ can be conveniently expressed as:

$$\mathcal{D}_{p,t} = \mathbb{E}_t^{\mathbb{Q}_T}\left[D_p(F_T, F_t)\right] = \frac{F_t^p}{p(p-1)}\mathbb{E}_t^{\mathbb{Q}_T}\left[e^{p\log F_T/F_t} - 1\right].$$

Define $\varphi_t(p, T-t) := \mathbb{E}_t^{\mathbb{Q}_T}\left[e^{p\log F_T/F_t} - 1\right]$ and assume that all requisite moments exist. Such a representation with the use of the moment-generating function of the log-return is widely popular in Finance, mostly because of the developments in the literature of affine jump-diffusion models (Duffie, Pan, and Singleton, 2000). The MGF completely characterizes the conditional distribution of the return and under the reasonable assumption that the return is independent of the level of the stock price prevailing at time t, only depends on the values of the (potentially latent) state variables. The total option costs/proceeds from settlement-replicating $D_\gamma^{n,p}$ are thus:

$$\mathcal{C}\left[D_p\right] = -\int_{t_0}^{t_n} \frac{F_s^{-p}}{p(p-1)} dF_s^p \varphi_s\left(p, T-s\right).$$

By applying Itō's formula to the function $f(x,y) = xy$, we can rewrite the integrator as:

$$dF_s^p\varphi_s\left(p,T-s\right) = \varphi_s\left(p,T-s\right)dF_s^p + F_s^p d\varphi_s\left(p,T-s\right) + \frac{1}{2}d\left[F^p, \varphi\left(p,T-s\right)\right]_s^c,$$

and plug it back into the cost expression.

$$\begin{aligned}\mathcal{C}\left[D_p\right] = &\overbrace{\frac{1}{p(p-1)}\left(\varphi_{t_0}(p,T-t_0) - \varphi_{t_n}(p,T-t_n)\right)}^{\text{Starting and ending swap rate}} \\ &- \underbrace{\int_{t_0}^{t_n}\frac{\varphi_s\left(p,T-s\right)}{p(p-1)}\frac{dF_s^p}{F_s^p}}_{\text{Rebalancing of the divergence position}} \underbrace{-\frac{1}{2}\int_{t_0}^{t_n}\frac{F_s^{-p}}{p(p-1)}d\left[F^p,\varphi\left(p,T-s\right)\right]_s^c}_{\text{Covariation between } F \text{ and state variables}}\end{aligned} \tag{1.3.13}$$

The costs of the dynamic option trading have three sources. First, establishing the initial static swap position, the first term in the above equation. Second, two types of costs associated with the rebalancing of the divergence position, the first of which corrects for changes of scaling required so that the portfolio payoff at maturity tracks the divergence of the log return, second, a term resulting from the fact that returns

[11] The trader *receives* the realised settlement payoff.

can be instantaneously correlated with changes in the price of divergence. The second term in equation (1.3.13) indicates that whenever a trader replicates a strategy that insures her against variance (i.e. she receives $D_\gamma^{n,p}$ at settlement), a positive return *decreases* the replication cost, while for a trader providing the payoff (i.e. short variance, paying $D_\gamma^{n,p}$ at settlement) a negative return generates proceeds from the option trading. Moreover, as noted before, the costs and proceeds from the rebalancing depend on the maturity of the options used for replicating $D_p(F_T, F_t)$; when replicating short-term payoffs, for example $D_\gamma^{n,p}$ over the course of a week, a trader might prefer to use options in the "weekly" CBOE series with the nearest maturity, rather than the standard monthly-calendar instruments, especially because longer-maturity instruments are exposed to risk long beyond the target maturity. The third term in (1.3.13) is related to the leverage effect: the most important driver of $\varphi_s(p, T-s)$ is the time-varying volatility.

In Appendix A.1.4 I derive the distributions of cost or proceeds of high frequency option replication of $D_\gamma^{n,p}$ and associated jump measures $S_{\gamma,J}^{n,p}$ and $Q_{\gamma,J}^{n,p}$ in the Black-Scholes model. In this setting the asset price variance is constant and the limiting divergence – non-stochastic, yet option trading costs are not zero. Similarly, even though in the Black-Scholes settings the price paths are almost surely continuous, the option strategy replicating zero at maturity does entail stochastic rebalancing flows. This is most instructive: in order to analyse the risk factors driving the premia associated with such dynamic option trading strategies, the Black-Scholes model remains an intuitive benchmark. The most important takeaway from this analysis is that the dynamic strategies are not *purely* exposed to jump risk, but they are *the only* model-free method of receiving the pure jump settlement payoffs. Thus I interpret average profits from such strategies as *premia for big risk* only because the settlement profits time series does indeed contain realisations of big, non-scalable risk.

The investor willing to obtain realized variation payoffs at settlement enters into an exchange of risks. The option cost component, described here in detail, exposes him to an interaction of return and price-of-variance risk. In essence, and investor gains from the incremental, single-period trade whenever the replicated power of the return is greater than the same return times the prevailing price of variance.

1.4 Feasibility of high-frequency trading at the CBOE

In recent years most high frequency studies of stock returns concentrated on the five minute frequency. It has been argued that at this frequency the return records are sufficiently dense to separate jump from Brownian innovations while keeping contamination from microstructure noise negligible. I investigate the results of dynamic option trading at two frequencies: in addition to the base five-minute frequency, I consider investors who rebalance every 1 hour during the trading day. I choose this frequency in order to compare some trading results with other studies of option payoffs, such as Muravyev and Ni, 2016. These two choices anchor my results in

a wider context and allow to form a bridge between the high-frequency econometrics and the asset pricing literatures.

Before I move to the description of how I implement the trading technology, I briefly describe the intra-day trading patterns in the market for short-maturity options at the CBOE. Then I discuss the technical details of obtaining sufficiently accurate replication, and finally I demonstrate the feasibility of accurate realized variation trading.

1.4.1 Short-maturity option trading at the CBOE

I analyse a subset of the complete set of CBOE's trade and quote records for S&P 500 index options, purchased from Market Data Express. The complete database holds records from January 2001 until December 2015, a total of 41 billion observations, 99.6% of which are quotes. My object of interest is the trading in weekly SPX options, between January 2011 and December 2015, over 261 option maturity cycles. Even though weeklies have been introduced in September 2007, they became sufficiently richly traded at the beginning of 2011. See Figure 1.4 for an illustration of how the number of strikes for which quotes are available, changed over time.

I aggregate the data to an hourly frequency, and a five-minute, using the "previous tick" rule, prevalent in the high-frequency literature. For each hour of records, for each option type (call or put), and for each strike, I pick the quotes with the latest time stamp available.[12] If multiple quotes share the same time stamp, I pick the lowest available ask price and the highest available bid price. I set the forward price of the underlying asset as the median of forward prices implied by put-call parity in the 5-minute interval. Within each period, I discard in-the-money options, as well as quotes for which $p_{ask}/p_{bid} > 5$. Furthermore, I discard quotes for which $log(K/F_t)/(\sigma_t^{IV}\sqrt{\tau_t})$ is smaller than -12 and greater than 8.[13] Finally, in order to alleviate the problem of using stale quotes, I use the procedure described in Appendix A.1.3 to ensure that at each time stamp the resulting option price system does not allow for static arbitrage.

1.4.2 Replication technology

At every time t_j the trader takes a position in the option market according to equation (1.3.11): she replicates (a linear combination of) payoffs $D_p(F_T, F_{t_j})$ and $\Psi_{p,k}(F_T, F_{t_j})$, say $M_{p,\gamma}(F_T, F_{t_j})$, generated by function $m : \mathbb{R}_+ \to \mathbb{R}$. If at each time options with expiry date T and strikes $K \in [0, \infty)$ were quoted in the market, she would take the following position in put ($P(K,T)$) and call ($C(K,T)$) options, following Carr and

[12] The quotes for many deep out of the money options are updated very infrequently because the market makers post very wide bid-ask spreads. The previous-tick rule means that often option quotes from many hours beforehand are used.

[13] This method of standardizing the strike range of analysed option data was used, among others, by Andersen, Fusari, and Todorov, 2015b

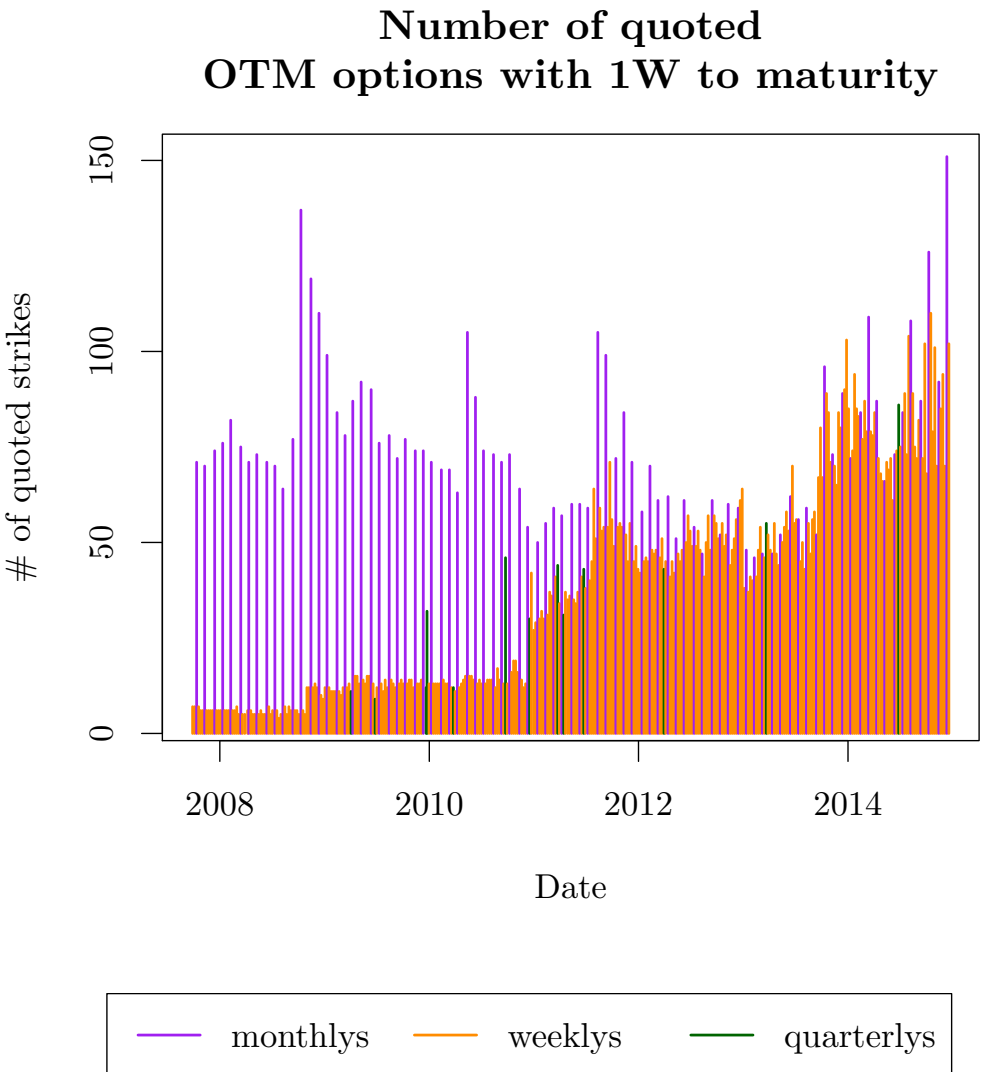


FIGURE 1.4: Number strikes at which out of the money options are quoted at the end of each day since the weekly options appear in the MDR sample, i.e. with maturity on 2007-09-28. Whenever a quarterly maturity is no more than 1 day away from the weekly maturity, and more quarterly options are quoted than weeklies, the number of quarterlys strikes is taken.

Madan:

$$\int_0^{F_i} P(K,T)m''(K)dK + \int_{F_i}^{\infty} C(K,T)m''(K)dK. \tag{1.4.1}$$

The most widely used method of forming an approximating portfolio resorts to a basic discretization of the integrals in the above formula, with options quoted for J strikes:

$$\sum_{k=1}^{\max k \text{ s.t. } K_k \leq F_{t_j}} P(K_k,T)m''(K_k)(K_k - K_{k-1}) + \sum_{\min k \text{ s.t. } K_k > F_{t_j}}^{J} C(K_k,T)m''(K_k)(K_k - K_{k-1}). \tag{1.4.2}$$

A similar technique with a special treatment of K_0 and K_J is used in the calculation of the VIX index (CBOE, 2000). The result is a portfolio of options with weights $w_k = m''(K_k)(K_k - K_{k-1})$. Such an approach has been considered sufficiently accurate in semi-static strategies. In dynamic option trading, such as settlement-replicating the payoff $D_\gamma^{n,p}$, it yields approximately 5% errors (MAPE). Recall that the higher order measures $S_\gamma^{n,p}$ and $Q_\gamma^{n,p}$ accumulate quantities orders of magnitude smaller than individual small time increments of $D_\gamma^{n,p}$.[14] Moreover, in, for example, a week of trading the replication has to be repeated up to 400 times: the accumulation of individual replication inaccuracies renders the method inadequate.

I improve on (1.4.2) by implementing the following optimisation problem. At time t_j the trader chooses an importance function $\eta : \mathbb{R}_+ \to R_+$ and then option portfolio weights $w \in \mathbb{R}^J$ which minimise the following objective:

$$\min_{w \in \mathbb{R}^J} \int_0^{\infty} \eta(F_T) \left(M_{p,\gamma}(F_T, F_{t_j}) - \sum_{k=1}^{J} w_k O(F_T, K_k) \right)^2 dF_T, \tag{1.4.3}$$

optionally s.t.

$$\text{sign}(w_k) = \text{sign}(m''(K_k)) \tag{1.4.4}$$

where

$$O(F_T, K_k) \equiv \begin{cases} \max(F_T - K_k, 0) & \text{if} \quad F_T > F_{t_j} \\ \max(K_k - F_T, 0) & \text{if} \quad F_T \leq F_{t_j} \end{cases}.$$

The most extreme available strikes are $0 < K_1 < K_J < \infty$, hence for many $M_{p,\gamma}$ criterion (1.4.3) will be unbounded unless η decreases sufficiently fast for $\log F_T \to \pm\infty$. The importance function η determines where on the support of at-maturity asset prices, payoff replication error is considered important. Natural choices for η are, for example $\eta(F_T) = f_0^{\mathbb{P}}(F_T|F_{t_0})$ and $\eta(F_T) = f_0^{\mathbb{Q}}(F_T|F_{t_0})$, the statistical and risk-neutral conditional distributions of F_T, if one assumes that sufficiently high moments of both exist. Likewise, simple truncation, i.e. $\eta(F_T) = \mathbf{1}_{\{K_0 - \varepsilon \leq F_T \leq K_J + \varepsilon\}}$, yields sufficiently accurate results, because the terminal asset price is almost never outside $[K_0, K_J]$. In practice, the choice of η is not crucial, however choices which

[14] $S_\gamma^{n,p}$ and $Q_\gamma^{n,p}$ can be though of as limits of first (second) differences of $D_{p_0 \pm h}$ around some p_0, scaled by h (h^2), and any inaccuracy in replicating D_p will be significantly augmented.

put more emphasis on accuracy around F_{t_0} than far away from that point are preferred, because they improve the relative replication error.

The problem in (1.4.3) is quadratic in w, irrespective of the choice of η. The details on the formulation and division into convex sub-problems are given in Appendix A.1.2. If no extra constraints such as (1.4.4) are required, (1.4.3) is equivalent to the L^2 projection of $M_{p,\gamma}(F_T, F_{t_i})$ on the space of piecewise-linear functions, weighted by $\eta(F_T)$, and the weights can also be obtained as solutions of a system of linear equations, if an adequate orthogonal basis is constructed from option payoffs. The convex quadratic programming formulation allows for very efficient numerical implementations.

This technology allows the investor who wishes to hedge a non-linear payoff in an incomplete market to choose their hedge optimally, and to minimise the expected discrepancy between the hedged quantity, and the received payoff.

1.5 Replication accuracy

I measure risk compensation by evaluating trading profits from big risk trading strategies. While Section 1.3 presented the theoretical underpinnings of trading realised variation measures, it remains to be shown that those strategies are indeed implementable. Only then, if settlement payoffs are closed to realised variation measures calculated at corresponding frequencies, can one make statements about premia for big risk.

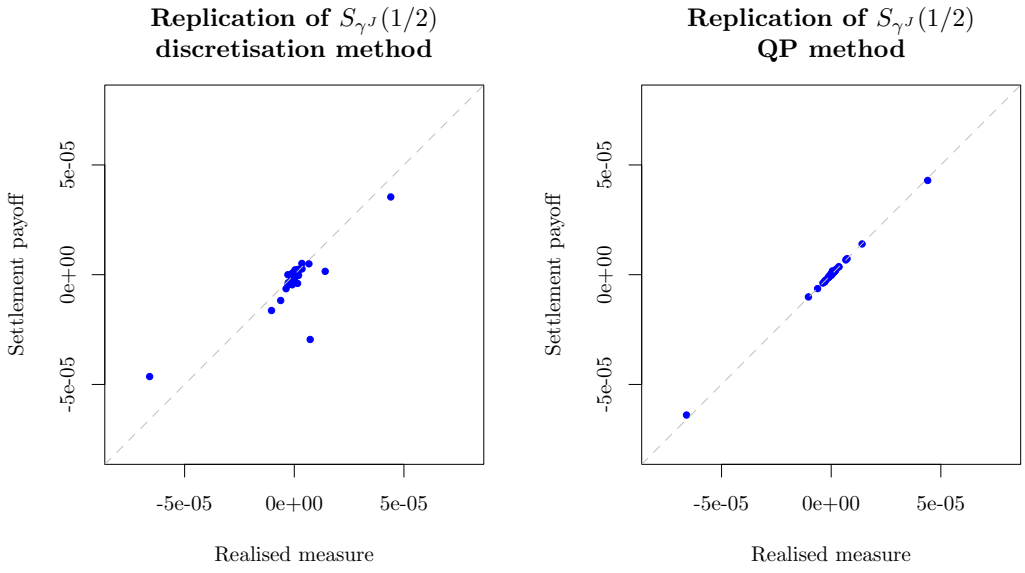


FIGURE 1.5: Comparison of replication accuracy of $S_{\gamma^J}(1/2)$ with the discretisation method (left panel) and the quadratic programming method (right panel).

Replication accuracy hinges on the QP-formulation of the problem in Section 1.4.2. Figure 1.5 presents the results of implementing the jump skewness trading strategy $S^{1/2}_{\gamma^J}(F)$ between 2004 and 2011 along the *monthly* settlement calendar with

nearest-to-maturity options.[15] Both panels present scatterplots of true realised skewness values, calculated from implied index forward prices, and corresponding replicated settlement payoffs. They should be aligned along the 45-degree line. In the left panel I present the simple discretization method (1.4.2). It is clearly inferior to the QP method (1.4.3), which offers almost perfect alignment in this period of trading.

I present detailed results for trading along the settlement calendar of weekly options, from 2010-12-31 to 2015-12-31. I consider two trading frequencies: with rebalancing every 1 hour, and every 5 minutes. The results are collected in Tables A.1 and A.2, respectively. A graphical presentation of the results is available in Figures 1.6 and 1.7. For all trading strategies, I compare the unconditional distributions of settlement payoffs and of the true realised measures, calculated from forward index prices. The Kolmogorov-Smirnov test cannot reject the null that both payoffs and realised measures come from the same distribution, but the p-values, reported in the rightmost columns of both tables, indicate that higher-order variation measures are indeed much harder to replicate accurately than divergence measures. This is clearly seen in the first six columns of both tables, where I report order statistics of both types of quantities. While in all cases the quartiles, mean and median are close, the case of in-sample minima for positive-valued measures requires attention. At times when $D_\gamma^{1/2}(F)$ and $Q_\gamma^{1/2}(F)$ are very low, the magnitude of replication error can dominate the payoff, and the result is *negative*. A closer inspection of the scatterplots in figures 1.6 and 1.7 indicates that the magnitude of the payoffs in such cases is small and if no bearing on the final results, except for the $Q_{\gamma^{JV}}^{1/2}(F)$ strategy, where in some cases the replication clearly fails. A final observation is that the magnitude of replication error is proportional to the variation of the strategy weighting process, γ.

Overall, I find that trading profits follow realised variation measure values closely enough to allow for an analysis of premia for big risk through the lens of dynamic option trading strategies. In certain cases, the observations (i.e. individual options) that cause significant replication error, can be identified and purged from the data. In the case of higher-order strategies, inaccurately replicated observations are often of outlying character, i.e. they are much greater in magnitude than the true realised variation measures. As such, they are removed from the sample before a closer look at the data.

1.6 Trading big risk

In this section I present empirical evidence about the compensation for big risk in the market for short-maturity S&P 500 index options. I reiterate that even though at times I made use of the semi-martingale continuous-time approach, mostly for the purpose of interpreting certain quantities, the evidence for the significance of pricing of big risks, presented below, does not depend on using that approach. The purely

[15] I do not present the general results for trading monthly options at this time.

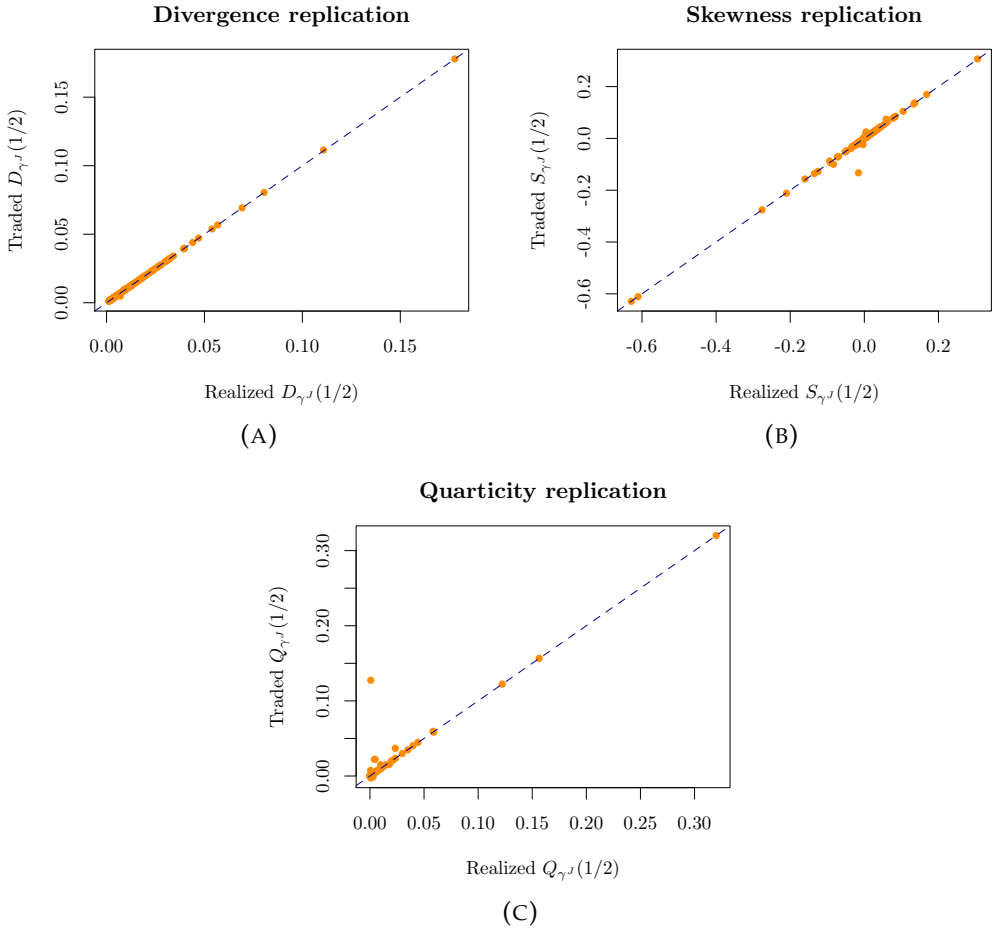
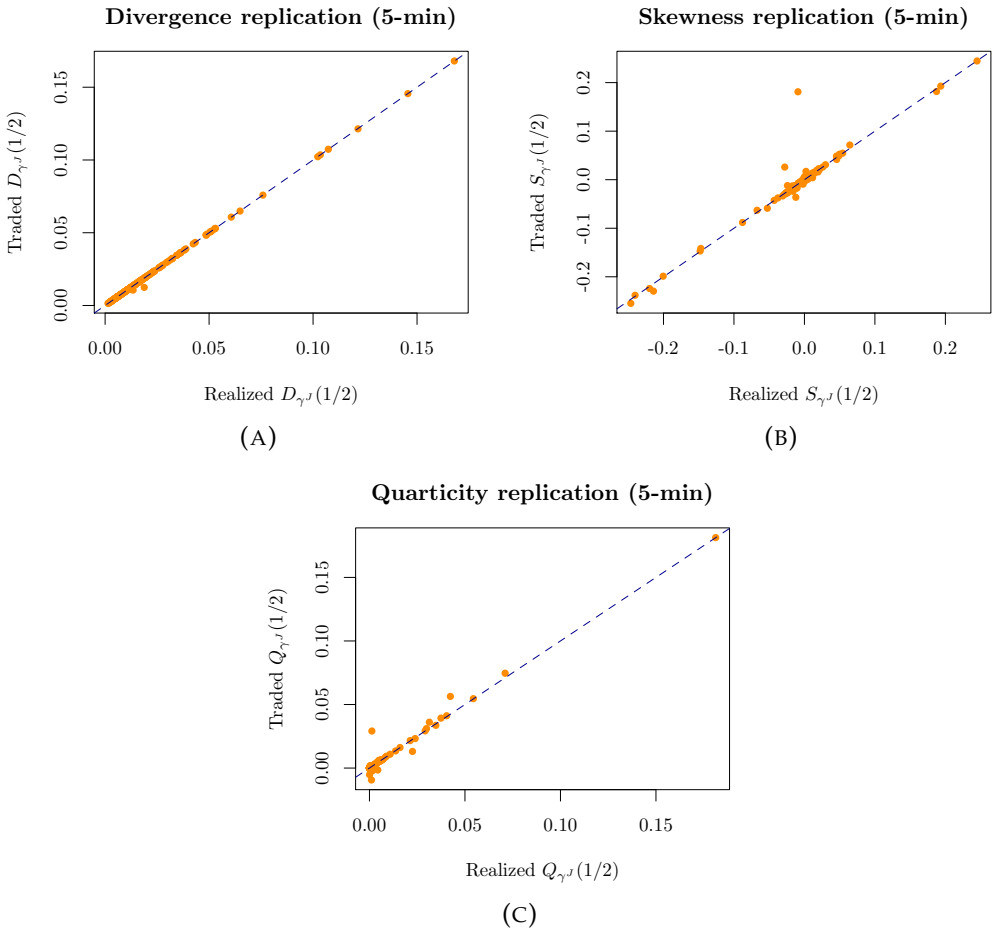


FIGURE 1.6: Scatterplots of realized divergence, skewness and quarticity with weighting γ^J, and dynamically replicated settlement payoffs, with weekly option settlement and 1-hour hedging interval. Trading sample from 2010-12-31 to 2015-12-24.



(A) (B) (C)

FIGURE 1.7: Scatterplots of realized divergence, skewness and quarticity with weighting γ^J and dynamically replicated settlement payoffs, with weekly option settlement and 5-minute hedging interval. Trading sample from 2010-12-31 to 2015-12-24.

algebraic transformation of the realized measure into a trading strategy in equation (1.3.11) is the foundation of all the results.

I consider a risk exposure to carry significant risk compensation, if the corresponding total payoff is persistently different from 0. While the theory of Finance defines a risk premium as $-cov(M, R)$, where M is the stochastic discount factor and R is the total payoff, M is unobservable and is not measurable in a purely non-parametric setting. The hurdle of non-zero profitability is a difficult one to clear from a statistical point of view, given the high variation of trading profits and settlement payoffs. In some cases, it is cleared even *after transaction costs*, i.e. when the bid-ask spread in the option market is taken into account. If a big risk strategy delivers non-zero profits after transaction costs, and the settlement payoff exhibits extreme variation, it is evidence that a premium is paid for the underlying exchange of risks.

I first briefly report on the statistical properties of settlement payoffs – or equivalently, realised divergence measures. Then I break down the analysis of trading profits into two parts. I report the unconditional descriptive statistics to provide an overview of the properties of engaging in big risk trading. Further on, I attempt to determine *when*, in what market conditions big risk trading is profitable. To that end, I accompany the standard (i.e. weekly) trading results with disaggregated evidence.

The analysis starts with trading options along the weekly settlement calendar. At the settlement of an option series,[16] I start the trading strategy in options expiring at the end of the following week,[17] and I rebalance the position at fixed time intervals, except for overnight trades (i.e. from 15^{15} to 08^{30} the following day, Chicago time) when hedging is not available.

The construction of my dynamic strategies allows for trading over arbitrary periods and at arbitrary frequencies. In the disaggregate analysis I take the following vantage points. First, I trade big risk from the opening to the closing of the Chicago option market; such an implementation offers exposure to big risk on single days, when active hedging is possible in many markets. Second, I trade close-to-open positions, which reduce to an opening trade around 15^{15} and a closing trade at 08^{30}, without intermittent option rebalancing nor trading in the forward contracts.[18] In the light of the findings in Muravyev and Ni, 2016, the daytime and night-time trading periods are significantly different in the option market to merit a more careful treatment. From the point of view of an actively trading investor, the overnight period is when she is forcefully exposed to big risk due to the inability to incrementally hedge the accruing variation.

Settlement payoffs from trading strategies with γ^J weighting are – by construction – of extremely small magnitude. Not only are they difficult to compare against

[16] Most SPX options are settled on Friday morning or afternoon.

[17] Unless there exists a contract which matures within one day of the default weekly option and is far more traded. This occurs for some quarterly option settlements in the early part of the sample.

[18] While CBOE now offers extended trading hours for option contracts, this is not the case from the start of my sample. Furthermore, outside regular trading hours liquidity virtually dries up.

each other, but also in the dimension of varying trading horizons. In order to facilitate the interpretation of trading profits, I rescale the results of trading $2S_\gamma^{1/2}(F)$ and $2Q_\gamma^{1/2}(F)$ by the sample averages of the prices of the corresponding static absolute variation swaps of corresponding maturity. Thus $S_\gamma^{1/2}(F_{0,T})$ is rescaled by

$$2\mathbb{E}_0^{\mathbb{Q}}\left[\left|S_{F_0^{-p}}(F)\right|\right], \tag{1.6.1}$$

and $Q_\gamma^{1/2}(F_{0,T})$ by

$$2\mathbb{E}_0^{\mathbb{Q}}\left[Q_{F_0^{-p}}(F)\right]. \tag{1.6.2}$$

The strategies $\left|S_{F_0^{-p}}(F)\right|$ and $Q_{F_0^{-p}}(F)$ are constructed by taking p-derivatives of $\gamma_s D_p$ with constant $\gamma = F_0^{-p}$, as in Definitions 1 and 2. They correspond to simple absolute third order variation, and kurtosis swaps with payoffs:

$$\frac{1}{2}\int_0^T \left|\ln(F_s/F_0)\frac{F_s^p}{F_0^p}\right|\sigma_s^2 ds + \sum_{0\le s\le T}\frac{\partial}{\partial p}\frac{F_{s-}^p}{F_0^p}D_p(F_s, F_{s-}) \tag{1.6.3}$$

$$\frac{1}{2}\int_0^T \ln^2(F_s/F_0)\frac{F_s^p}{F_0^p}\sigma_s^2 ds + \sum_{0\le s\le T}\frac{\partial^2}{\partial p^2}\frac{F_{s-}^p}{F_0^p}D_p(F_s, F_{s-}). \tag{1.6.4}$$

These payoffs arise naturally as a generalisation of Simple Variance Swaps (Martin, 2012) to higher order measures, and are studied, among others, in Schneider and Trojani, 2015b. They can be seen as *weighted* realised variance measures, with more weight put on variation far away from the forward price at inception due to the factor $\ln^k(F_s/F_0)F_s^p/F_0^p$. Rescaling the payoffs by sample averages of these quantities renders the results easier to interpret. $\mathbb{E}_0^{\mathbb{Q}}\left[\left|S_{F_0^{-p}}(F)\right|\right]$ is typically about $1/8$th of the value of a variance swap. Thus trading results are given in *multiples* of the prices of static higher order variation swaps.

The extreme nature of both settlement and total payoffs makes the estimation of their distributions' location parameter challenging. When evaluating the profitability of the trading strategies, I turn not only to the averages of the payoffs, but also to medians. I form confidence intervals for both estimates of location by bootstrap methods.

1.6.1 Weekly trading

I implement the weekly trading strategies at two hedging frequencies: with hourly and five-minute rebalancing. The summary statistics of realized variation measures, aggregate option costs, and trading profits are collected in Tables A.3 and A.4, respectively. I first briefly describe the time-series properties of the settlement payoffs, and then move to reporting on risk compensation.

Time-series properties of settlement payoffs

An overview of the time series of settlement payoffs is available in Figures 1.8 (for hourly hedging) and 1.9 (for five-minute hedging). In both plots, the top-left panel represents realised divergence, which does not have a big risk interpretation, but is given as a reference whose time-series properties are well-known, an analogue of realised variance. The top right figure plots the time series for realized big skewness $S_{\gamma^J(1/2)}$, while the bottom plot presents realized big quarticity, $Q_{\gamma^J(1/2)}$. This layout and colour coding – orange for $D_{\gamma^J(1/2)}$, blue for $Q_{\gamma^J(1/2)}$ and pink for $Q_{\gamma^J(1/2)}$ are retained through the remainder of the paper.

It is immediately visible that the big risk measures have time series properties that are very different from those of realised divergence. As the hedging interval shrinks to 5 minutes, the clustering of large movements becomes less pronounced, and two big risk periods stand out. These periods are of different character. The first, in the second half of 2011, coincides with a prolonged period of raised uncertainty due to the European debt crisis. Over the course of six months realisations of big risk are of greater magnitude than in all of the ensuing sample, except for the second stand-out moment. The latter is related to the Asian markets sell-off in August 2015, and the extreme realisation of all jump variation measures happens on 2015-08-24, the Monday when Asian markets dropped by 7%, and in the US the S&P500 was reflecting that movement in early trading. This event, also known as the "mini flash-crash" did not cause a prolonged period of large market movements.

At the higher hedging frequency the core properties of the time series of settlement payoffs are the same as in Section 1.6.1. The key difference – as seen in Figure 1.9 – . The higher hedging frequency allows to capture the dip and recovery in the price of the S&P500 index, which lasted around 30 minutes. The replicated skewness is strongly negative, which implies that the downward movement in the "mini flash-crash" was much more abrupt than the following recovery. The magnitude of settlement payoffs in other periods is lower, but similar to levels recorded at the one-hour trading frequency.

Contrary to measures of realised volatility, the measures of big risk variation are neither persistent nor predictable, which can be seen in Figure 1.10. While the reported plots present results for weekly trading at the 1-hour frequency, the results at shorter trading horizons and higher frequencies are not unlike those presented. These results are in line with other studies of jump variation.

Weekly trading profits

Trading profits for both trading frequencies are summarised in Tables A.3 and A.4. Table A.3 contains summary statistics for the hourly hedging frequency, while tables A.4 contain summary statistics for the five-minute hedging frequency. In each table, data on trading $D^p_\gamma(F)$ are contained in panel A, on $S^{1/2}_\gamma(F)$ in panel B, and on $Q^{1/2}_\gamma(F)$ in Panel C.

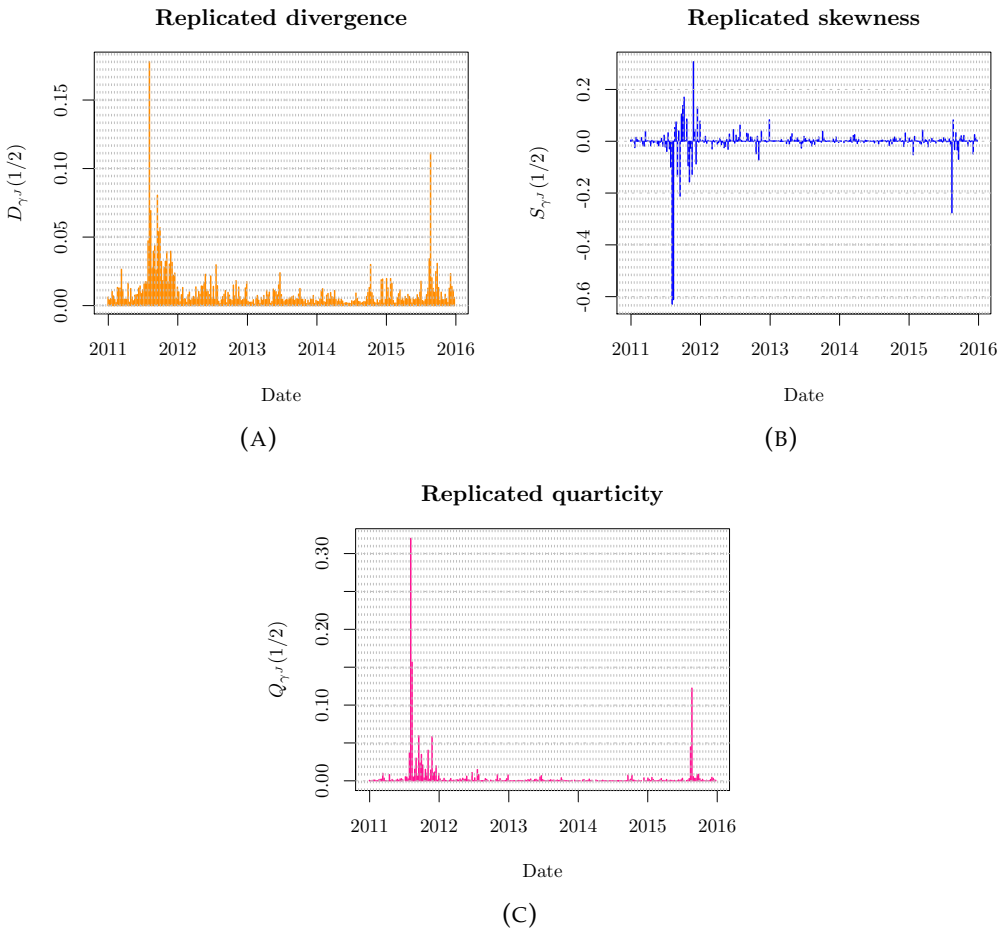


FIGURE 1.8: Time series plots of replicated divergence, skewness and quarticity trading strategies with weighting γ^J, with weekly option settlement and 1-hour hedging frequency. Trading sample from 2010-12-31 to 2015-12-24.

The summary statistics are reported at both mid-quote and bid-ask option cost calculation. The mid-quote profits, denoted $\mathcal{P}[\cdot]$, are reported for *long* strategies, in which the investor receives the settlement payoff at maturity. The transaction-cost profits, denoted $\mathcal{T}[\cdot]$, are reported for *long* strategies if average $\mathcal{P}[\cdot]$ for the corresponding strategy is positive, and from *short* strategies otherwise. This convention makes evaluating the final profitability of the strategies easier. Thus, for example, in panel A of Table A.3, the average profits for trading divergence at 1-hour frequency are negative at mid-prices, while the profits after transaction costs reported in the following line, are of a short strategy, and are positive; furthermore, the bootstrapped confidence interval does not contain 0. Whenever the after-transaction-cost profits $\mathcal{T}[\cdot]$ are *negative*, it implies that the transaction costs render profiting from the strategy infeasible in the CBOE option market.

First, I investigate the γ^J pure-jump strategies and also comment on divergence trading profits. At both trading frequencies I find that mid-price profits from trading big skewness, $\mathcal{P}[S_\gamma^{1/2}(F)]$ are, on average, positive, and of similar size. The average profits are between 0.14 (median) and 0.27 of the corresponding absolute third order variation swap rates. The standard deviation of the profits is much higher than in

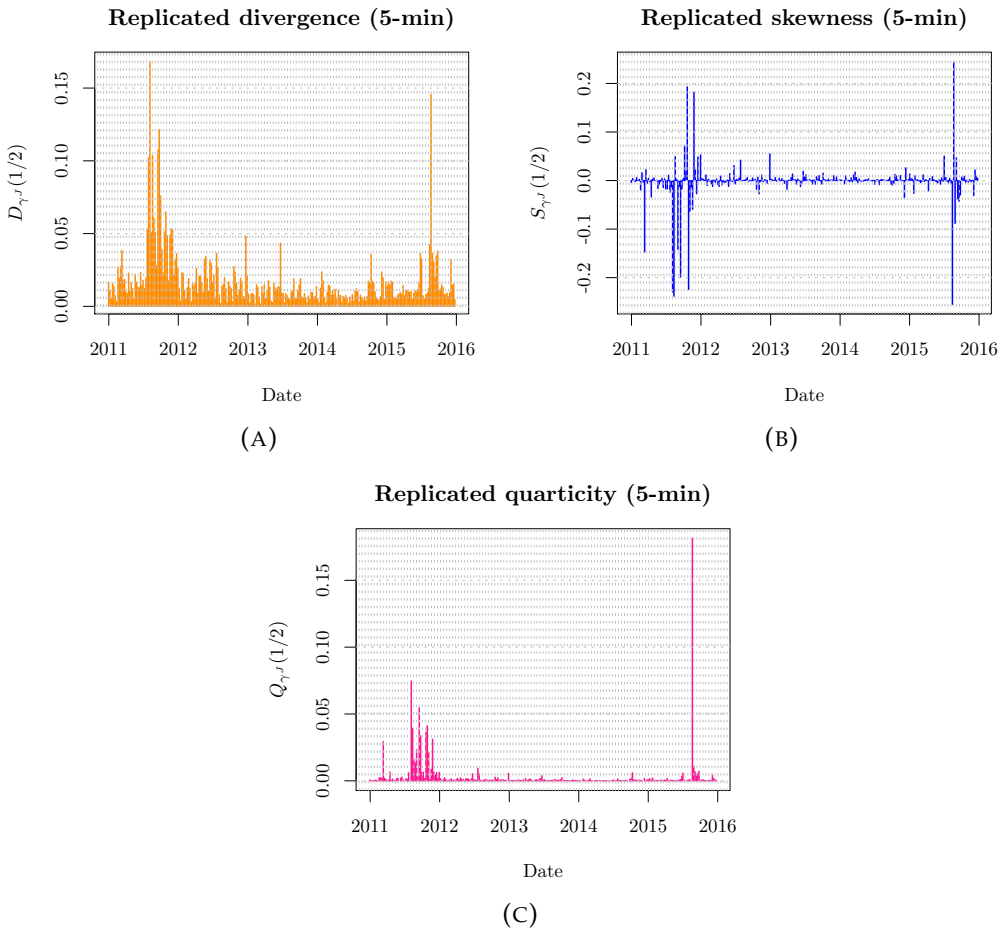


FIGURE 1.9: Time series plots of replicated divergence, skewness and quarticity trading strategies with weighting γ^J, with weekly option settlement and 5-minute hedging frequency. Trading sample from 2010-12-31 to 2015-12-24.

the case of divergence (variance) trading. Furthermore, the skewness strategies offer a very different higher moment properties, than divergence strategies (exposed to non-directional, small risk). Average $\mathcal{P}[S_\gamma^{1/2}(F)]$ are significantly different from 0 at both the 1-hour trading frequency, and at the 5-minute trading frequency. In both cases, median profits are of the same size, positive, and significantly different from 0. After transaction costs are taken into account, in both cases the confidence intervals for average trading profits contain 0. It is not the case for the median profits, however.

Trading big quarticity at both frequencies is "insurance", similarly to the results reported in the literature on trading variance over longer horizons (e.g. Carr and Wu, 2009b). $\mathcal{P}[Q_\gamma^{1/2}(F)]$ is significantly negative at both frequencies. Including transaction costs in the calculation erases all potential profitability at the high trading frequency, but allows for positive average profits from a short strategy at the 1-hour frequency; the latter quantity, however, is not statistically greater than 0. Median estimates of location indicate in both cases that mid-price average profits are indeed negative, and the median profits from the short strategy pass the statistical significance hurdle after transaction costs.

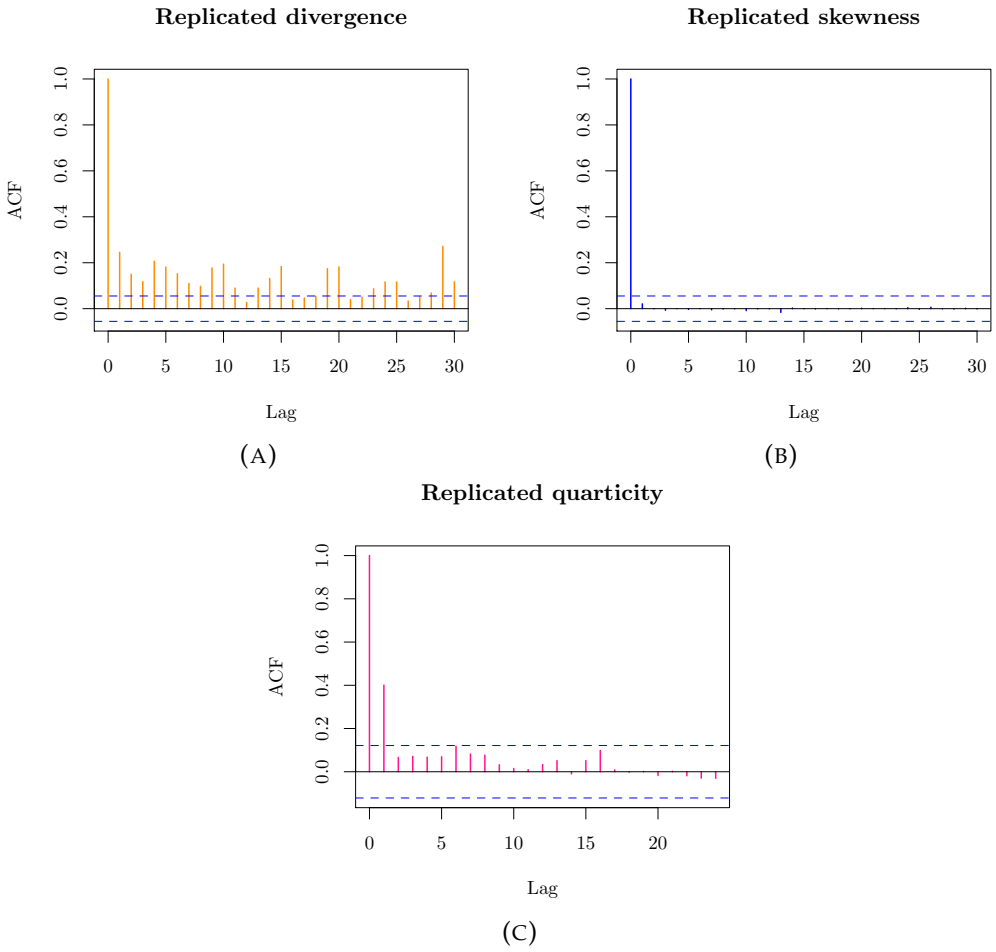


FIGURE 1.10: Autocorrelation function plots of traded divergence, skewness and quarticity with weighting γ^J (left-hand plots) and γ^{JV}. Trading sample from 2010-12-31 to 2015-12-24.

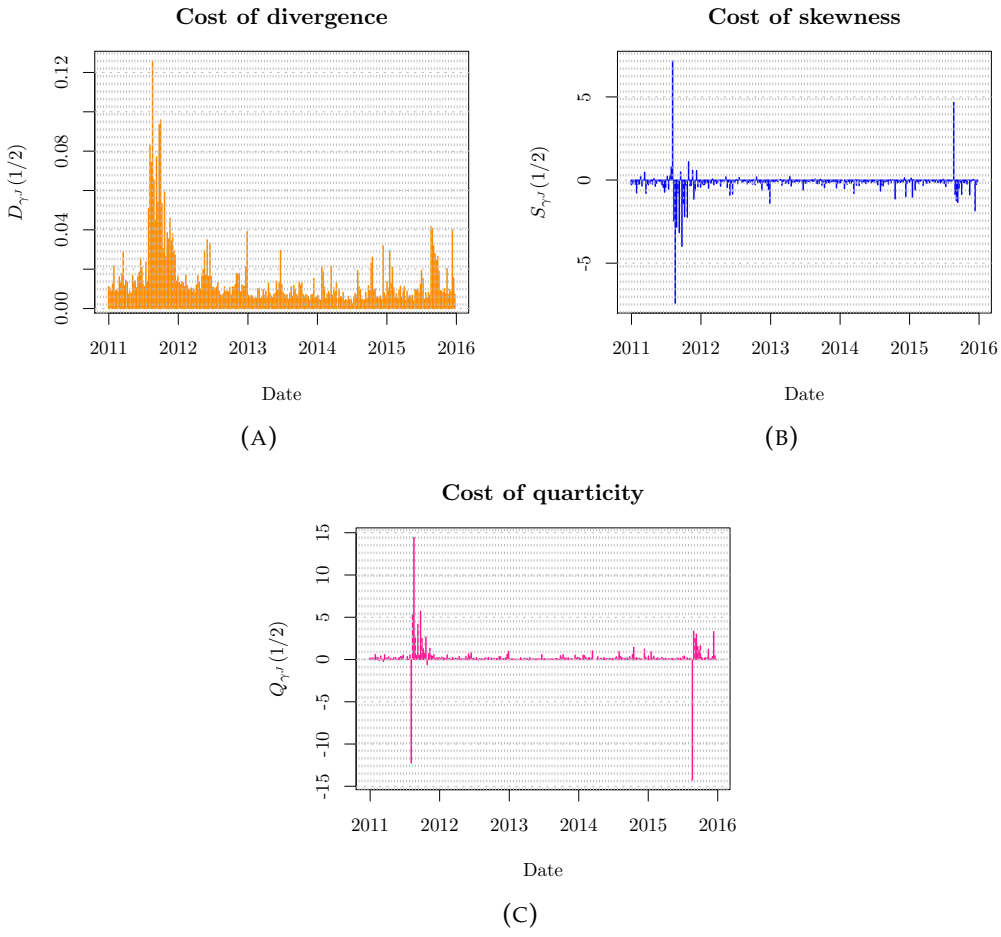


FIGURE 1.11: Time series plots of cost of replicating divergence, skewness and quarticity trading strategies with weighting γ^J, with weekly option settlement and 1-hour hedging frequency. Trading sample from 2010-12-31 to 2015-12-24.

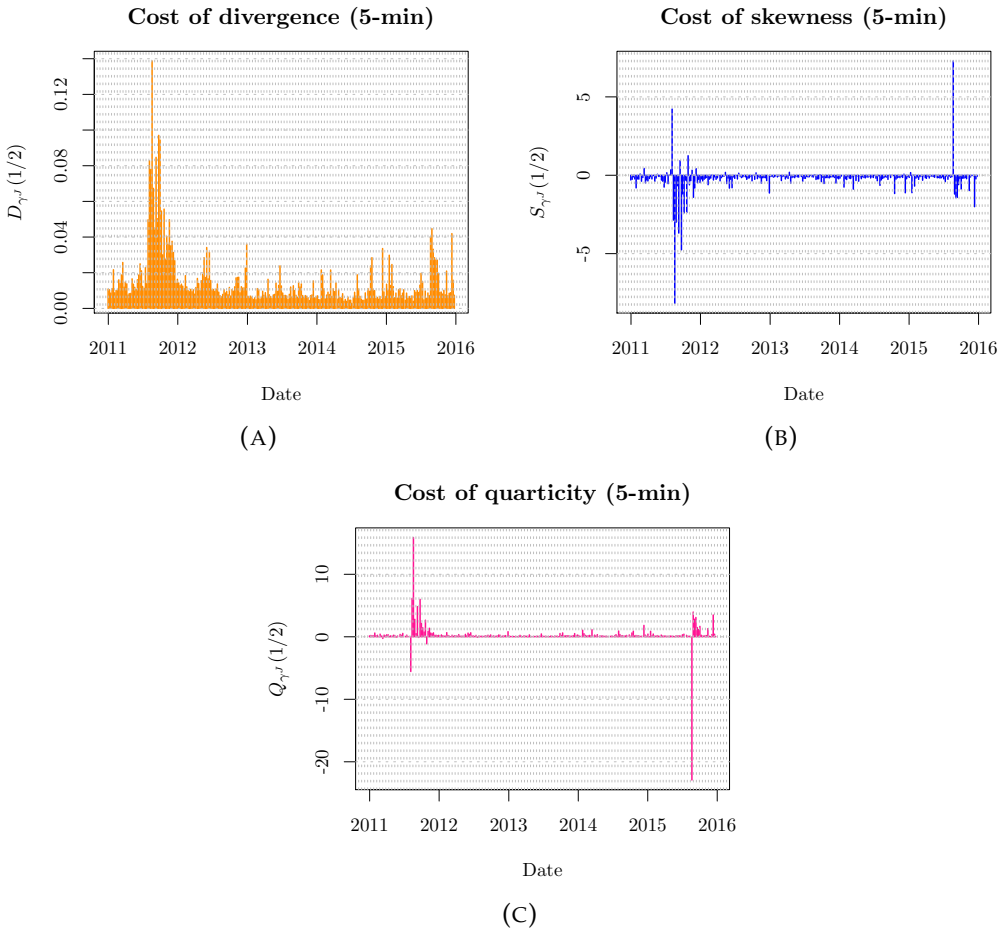


FIGURE 1.12: Time series plots of cost of replicating divergence, skewness and quarticity trading strategies with weighting γ^J, with weekly option settlement and five minute hedging frequency. Trading sample from 2010-12-31 to 2015-12-24.

An interesting finding is that the sign of the profits from trading divergence with the use of short-maturity options, changes when the frequency increases to 5-minutes. That is, the short-maturity variance risk premium in the period under consideration is positive, if hedging is frequent enough. This result is not driven by rebalancing in the option market, because it prevails if the $D^0_\gamma(F)$ strategy is considered, the VIX^2-based variance swap.[19] The profits from that strategy are virtually identical, as can be seen in the lower part of panel A in Table A.4. As a matter of fact, the higher divergence trading profits at the higher frequency are a result of an increase in the settlement payoff. Simply put, it is due to a more precise measurement of realized return divergence, which is fully reflected in the payoff.

To the contrary, the average and median quarticity strategy profits do not change sign with a change in the hedging frequency. At a high frequency and with short-maturity options, that price events only in the nearest future, investors are compensated for taking on extreme risks such as in the quarticity and skewness strategies. Compensation for the magnitude of small risks, which are the main driver of realized return divergence, is positive in the sample and horizon under consideration.

The time series of trading profits can be inspected in Figures 1.13 and 1.14. When viewed alongside Figures 1.11 and 1.12, which present the aggregate option costs defined in equation (1.2.10). The fundamental observation is that total payoffs for the big risk strategies are mostly generated by aggregate option costs, and not by the settlement payoffs. In the case of skewness trading, the settlement payoff is of greater magnitude than option rebalancing costs under 7% of the time, while for quarticity there are only 2 such instances in weekly trading. This is hardly surprising in light of the analysis in Section 1.3.2: the rebalancing costs accrue proportionally to return-weighted price of *divergence*, which is a magnitude greater than skewness or quarticity before rescaling is considered. This observation is important: it renders how difficult it is to isolate big risk in extant financial markets. That some of these strategies are profitable, on average, at mid-prices is indicative enough of the fact that big risk is a distinct feature of the S&P500 index option market.

1.6.2 Disaggregated trading

One of the most interesting properties of the trading strategies I consider is that they allow the investor to accrue the settlement payoff over arbitrary periods. I exploit this feature in order to study the differences in trading small and big risk during daytime (market opening hours), and overnight (from the time the market closes, to the time it reopens). This investigation is directly motivated by Muravyev and Ni, 2016, who found a striking pattern in option returns across all maturities (longer than two weeks) and wide strike ranges. Returns on delta-hedged option positions held from market open to market close are positive, while overnight returns are negative and so big in magnitude that overall returns on options are negative. I

[19] Recall that this strategy corresponds to $\gamma^J_t := F^0_t \equiv 1$, so it belongs to the family of semi-static strategies.

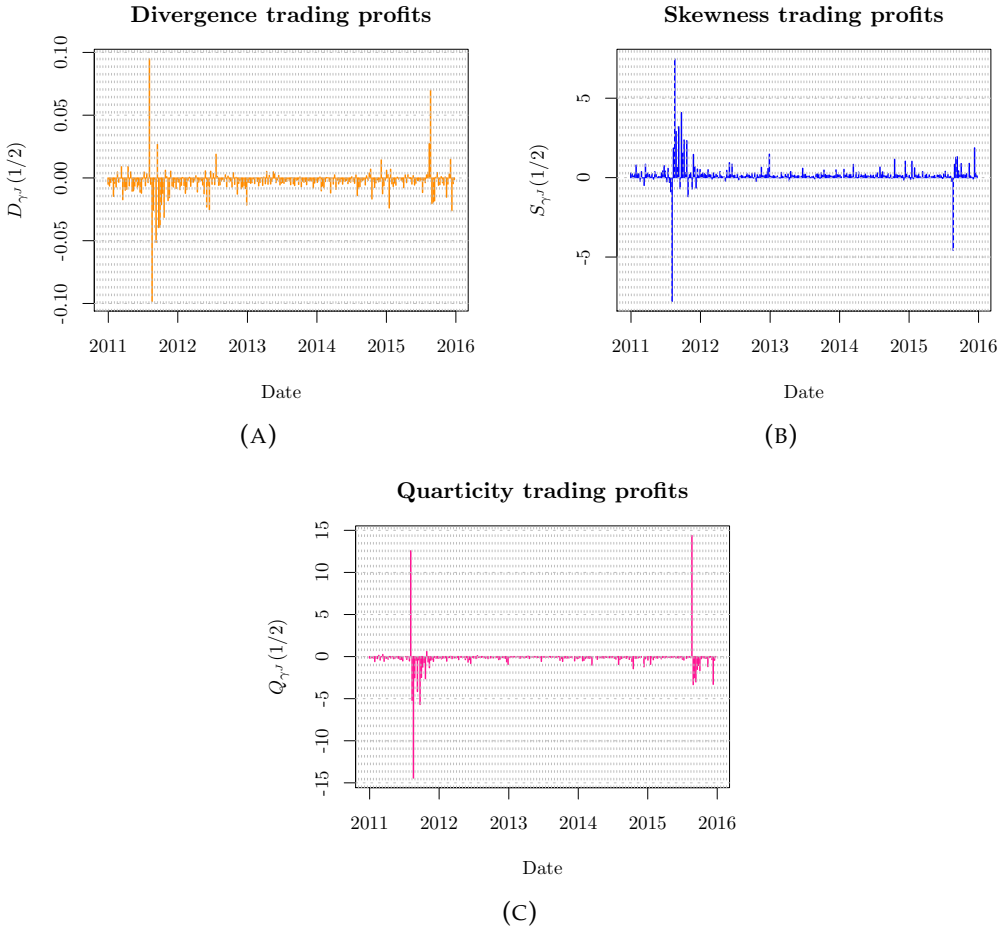


FIGURE 1.13: Time series plots of profits from replicating divergence, skewness and quarticity trading strategies with weighting γ^J, with weekly option settlement and one-hmy hedging frequency. Trading sample from 2010-12-31 to 2015-12-24.

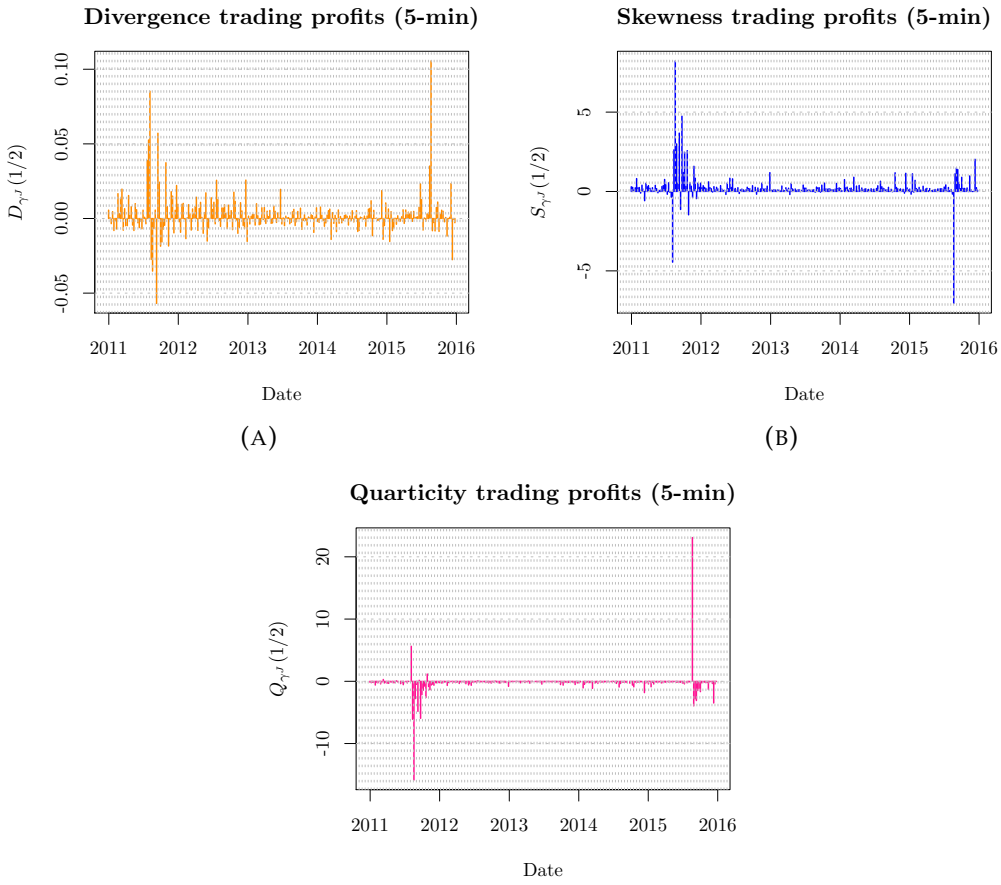


FIGURE 1.14: Time series plots of profits from replicating divergence, skewness and quarticity trading strategies with weighting γ^J, with weekly option settlement and five-minute hedging frequency. Trading sample from 2010-12-31 to 2015-12-24.

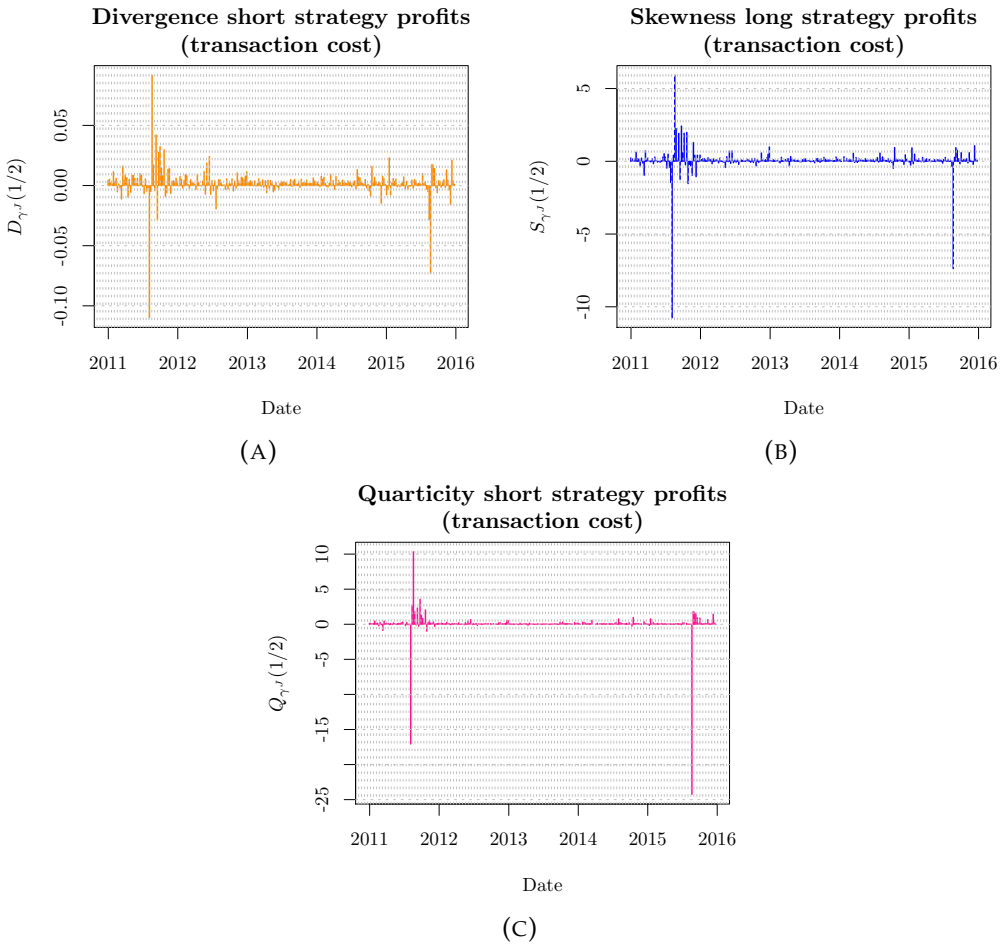


FIGURE 1.15: Time series plots of profits (including transaction costs) from replicating divergence, skewness and quarticity trading strategies with weighting γ^J, with weekly option settlement and one-hour hedging frequency. Trading sample from 2010-12-31 to 2015-12-24.

present results from trading during the day and overnight separately in Tables A.5 and A.6.

Time series properties of settlement payoffs

The general properties of the time series of daytime and night-time settlement payoffs are identical with those already reported for weekly trading. It is most instructive, however, to compare the two time series side by side (Figure 1.16). In the case of non-directional strategies, the overnight variation captured by the settlement payoffs is of significantly smaller magnitude than variation during daytime. In the extreme case of trading $Q_\gamma(1/2)$, all payoffs except for the "mini-flash-crash" are virtually equal to 0. In the case of a directional trade – big skewness – the daytime variation is smaller than overnight variation due to the fact that during daytime, frequent measurement will allow for cancelling out potential upwards and downwards jump movements.

Daily and overnight trading profits

In daytime trading, the average mid-price profits for trading big skewness are negative, but not significantly different from 0. The median mid-price profits remain positive. Transaction costs are, however, high enough to erase all potential profitability. Daytime average mid-price quarticity profits are positive, however median mid-price quarticity profits are not significantly greater from 0. Both average and median post-transaction cost quarticity profits are negative. These results imply that during daytime trading there is scant unconditional evidence for compensation for big risk.

The compensation seems to be present, however, in overnight trading. Average mid-price profits for trading big skewness are positive, as are median mid-price profits. Both location estimates are significantly greater than 0. The bid-ask spread, however, seems an obstacle to actually cash in on those payoffs – average transaction-price profits from a long strategy are negative, but median profits remain significantly greater than 0. Trading big quarticity is – unlike during the day – "insurance", in the sense that it carries a negative premium overnight. Both average and median profits are significantly negative, and the compensation is high enough so that the average profits remain positive after transaction costs are taken into account. When one takes into account how small overnight quarticity is on all days except around 2015-08-24, it is clear that the strategy is a hedge against really crippling events.

Overnight profits from trading big risk are also much greater (in absolute value) than daytime profits. After scaling by the $1/\tau^k$ factor, they are also greater than aggregate week-time profits, which is another indicator that compensation for big risks is present – unconditionally – in overnight trading.

The data on the divergence trading strategy $D_\gamma(1/2)$ offers additional insights into the question of the sign of the divergence premium. In Section 1.6.1 I noted that at the 5-minute trading frequency divergence profits over the weekly trading horizon are positive, on average. After disaggregation into daytime and night-time profits, a clear patter emerges: during daytime, the high-frequency divergence trade earns a positive risk premium, but overnight the premium is negative. The daytime premium is, however, of greater magnitude than the negative night-time premium.

The difference in premia when trading at various frequencies is, as noted before, mostly due to obtaining more precise "measurements" of the realized measures in the settlement payoffs. It remains an open question as to what hedging frequencies investors have in mind when engaging in such trades. One hypothesis is that there are many kinds of agents in the financial market, heterogeneous in their decision horizons. In that case the explanation of the option prices being "insufficiently high" to bring around a negative divergence risk premium – compatible with most to-date evidence – would lie in the interplay between the two groups of traders in the financial market.

Overall, I conclude that the compensation for big risk strategies in extant financial markets comes mostly through overnight trading. Thus, it is related to the fact that in these periods agents cannot engage in actively hedging their positions.

Conditioning

A careful look at the time-series plots of realised big variation measures (Figure 1.9) and trading profits (Figure 1.14 indicates that trading big risk might become highly profitable *after* such an event is actually observed. Such outcomes prevail in weekly trading for the big skewness strategy $S^{1/2}_{\gamma^J}(F)]$ and the short big quarticity strategy $Q^{1/2}_{\gamma^J}(F)]$ (Figure 1.14b and 1.14c, respectively). In disaggregate trading results more than 1200 daytime and overnight observations are available for a more detailed analysis.

I present more evidence backing up this claim in Figure 1.17. In all four panels, median mid-price profits from big risk trading strategies are plotted, conditional on the realisations of the previous period's corresponding realised big variation measure. For daytime trading (in the left two panels), the preceding period is from the previous market close to the current day's market opening. For night-time trading, the preceding period is from the current day's market opening, until market closing, when the position is taken. The top two plots present results for trading big skewness, while the lower two – big quarticity.

During daytime trading, there is no evidence for the dependence between median profits and realised variation. The patterns are striking for overnight trading. For both strategies, there exists a non-linear relation between the magnitude of preceding period's realised big variation, and the current period's trading profits. Compensation for skewness increases dramatically after a big risk event in the 10^{th} decile,

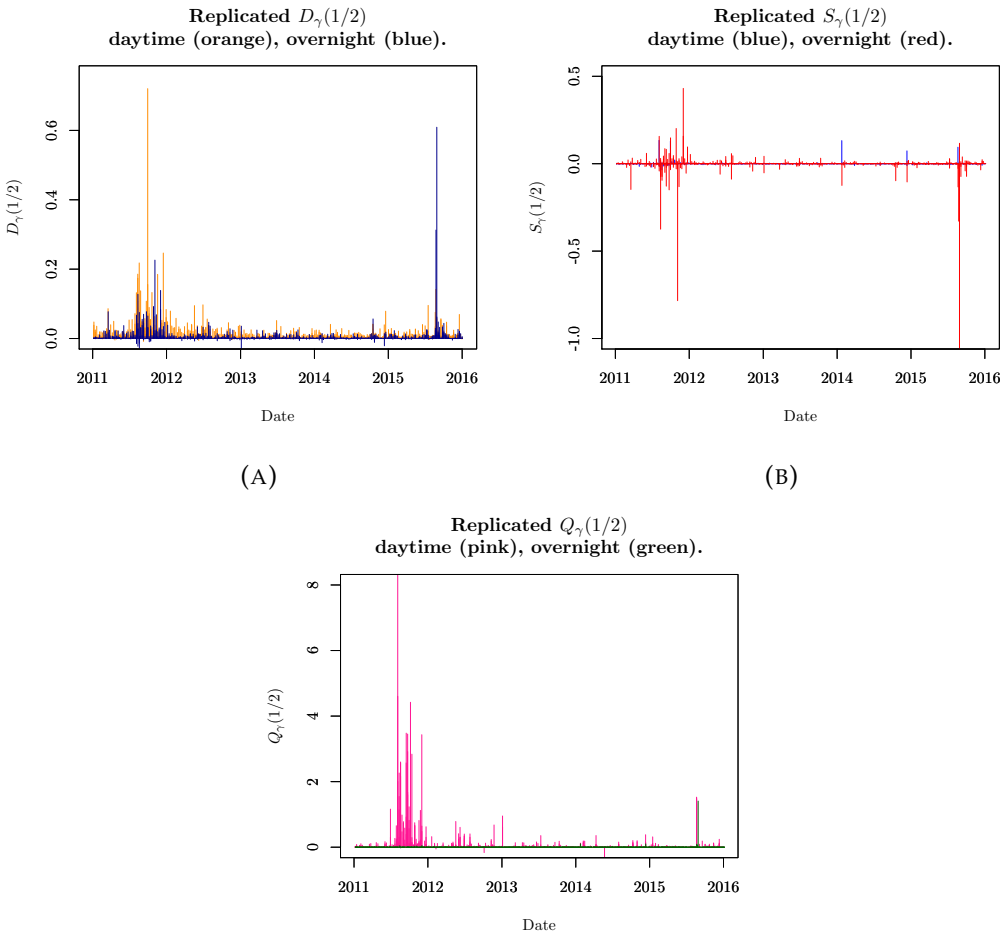


FIGURE 1.16: Time series plots of profits from replicating divergence, skewness and quarticity trading strategies with weighting γ^J, with daily and overnight option settlement and five-minute hedging frequency. Trading sample from 2010-12-31 to 2015-12-24.

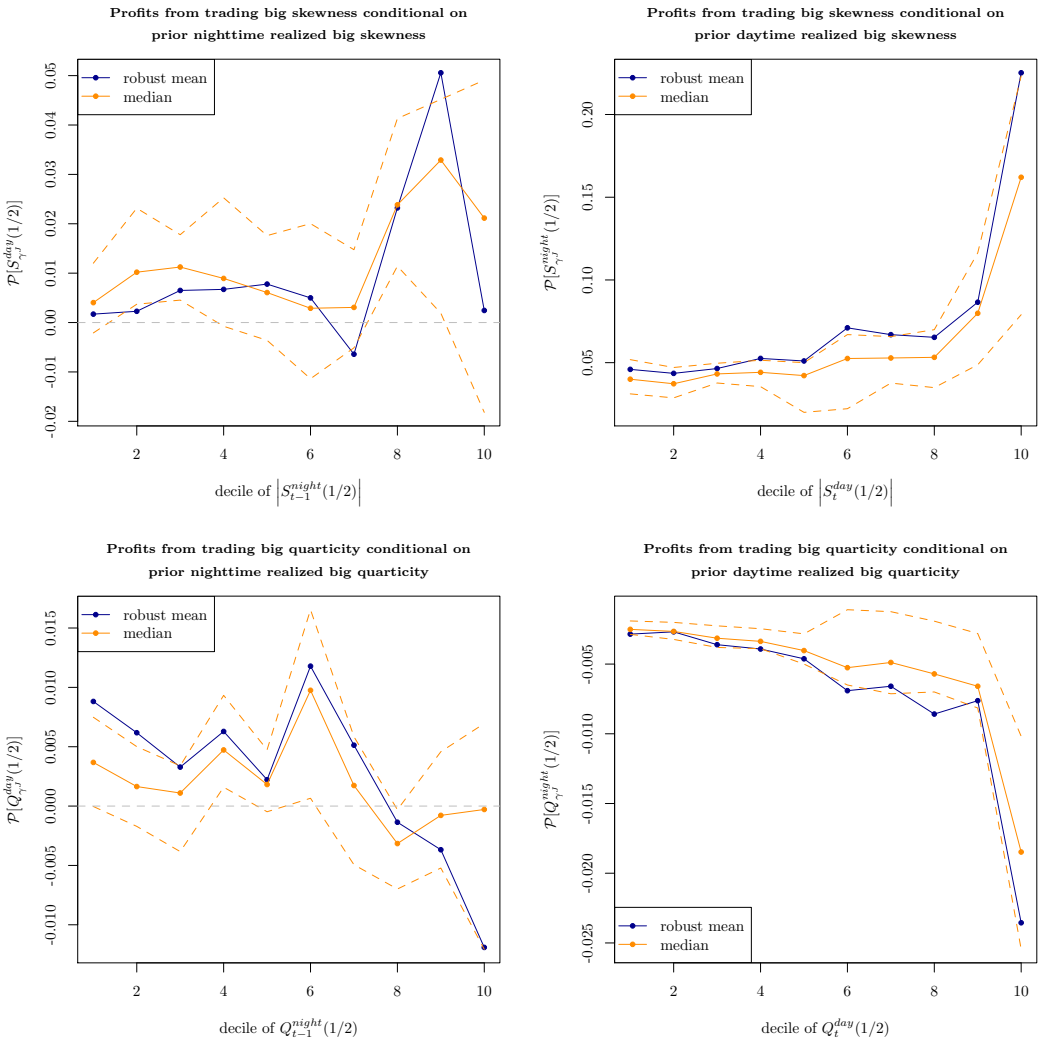


FIGURE 1.17: Profits from trading big skewness and big quarticity with weekly options, conditionally on realised big variation over the preceding period. For daytime trading, the preceding period is from previous closing until opening of markets on day of trading. For nighttime trading, the preceding period is from the day's market opening until market closing on night of trading. During the day hedging is at 5-minute frequency.

and so does the price of insurance against quarticity. Both these strategies load heavily on out of the money options. The increase in skewness compensation (or in the price of insurance against quarticity) reflects an increase in put prices after a truly large market movement. This result is consistent with evidence presented by Andersen, Fusari, and Todorov, 2015b, who parametrically model option prices and postulate the existence of a "tail risk" factor, which mostly influences deep out of the money put option prices. This factor is very persistent under the risk-neutral measure, but such extreme jump movements very rarely occur under the statistical probability measure.

1.7 Conclusions

In this work I introduced a general concept of dynamic option and forward trading, which opens new avenues for the asset pricing literature. The trading strategies' settlement payoffs can exactly hedge a family of realized return variation measures, that until now could only be calculated from asset price data. Rendering such specialized payoffs tradable allows researchers to put a price on precise concepts of asset price variation, including the possibility of separating big (jump) risk from small (diffusive). The strategies are based on active rebalancing of positions in currently known synthetic derivatives similar to variance swaps.

The feasibility of implementing such trading strategies in the option data hinges on using option portfolio formation methods that improve upon simple discretization, as has been used by practitioners and researchers alike. I show how to implement a conceptually simple and computationally fast method for obtaining optimal replicating option portfolios in incomplete option markets.

I expect the strategies to be useful for expanding the knowledge about the contribution of higher-order risks to the cross section of stock returns. The excess payoffs of the strategies, that is the difference between the settlement payoff and the aggregate option costs, reflect accumulated information about the investors' perception of higher order risk, and their reward for engaging into trading it. In hereby unreported results I find that the excess payoffs are not related to currently fundamental risk factors, such as Market, Size, Value (Fama and French, 1993) and Momentum (Carhart, 1997). Their response to investor perception of the price of variance can be useful in capturing yet unexplained cross-sectional expected return puzzles, especially in fields such as returns on hedge fund investments, which have relatively disorderly higher moment properties.

The properties of the trading strategies allow me to provide evidence about compensation for divergence (variance) and jump risk on an aggregate (over the weekly settlement horizon) and disaggregate (separately for daily and overnight trading) basis. There is a number of important findings. First, the divergence (variance) risk premium in weekly trading is negative at the hourly trading frequency, and positive at the very high 5-minute trading frequency. The change comes from the fact that in the latter case the realized measure is higher: an increased measurement frequency better captures small Brownian variation, which otherwise washes away. This effect for the divergence premium prevails also for the VIX^2-based divergence strategy, which does not involve active option between opening and closing the position. The finding complements a large body of literature about the term structure of variance risk premia. Almost all of this literature is based on longer-maturity contracts, and reports a negative risk premium for index variance.

The premia for big risk do not change that dramatically with an increase in the hedging frequency. I consider two contracts, paying of "big skewness" and "big

quarticity". The fundamental difference between them is that the former is directional, i.e. a long position brings losses if negative jumps occur, while the latter is convex and similar to a variance swap: large positive and negative jumps count positively towards the settlement payoff. The compensation for big skewness risk is positive in the market for very short maturity options, albeit the profits are smaller than in the case of divergence trading after similar financing capacity of the investor is considered. The compensation for big quarticity risk is negative, i.e. the strategy serves as a hedge against the magnitude of jumps. The profitability of the strategy on a weekly basis is, however very low, especially with five-minute hedging. If transaction costs (bid-ask spread in the option market) are taken into account, the average profits from the strategies are not statistically different from 0 anymore. It is worth noting, however that median profits are different from zero, and that transaction costs in the CBOE SPX option market are considered high among practitioners. Overall, I find evidence that investors require a premium for taking on directional jump risk, and are willing to pay to hedge against the magnitude of jump risk. The premium should be understood as this for an exchange of two risk profiles, however: of the pure realized variation measure for appropriately weighted price of divergence (variance).

The disaggregate analysis brings further insights into the nature of dynamic option trading. I report three important findings. One is about trading the divergence of stock returns, i.e. a contract exposed to both small and big risk. The latter two are about pure big risk trading. First, in trading the divergence of returns, I obtain a result similar in spirit to Muravyev and Ni: excess payoffs from trading divergence are positive during the day, and negative in the overnight periods. The daytime profits are, on average, much higher than night-time losses associated with a long position. The difference in excess payoffs from trading return divergence can mostly be attributed to the fact that overnight return divergence is on average smaller than daytime return divergence. Second, in big risk trading the compensation for risk is earned mostly in the overnight period. Profits from trading big skewness during daytime are not significantly different from zero, but they are considerable during night-time: profits from trading big skewness are of the same magnitude as profits from trading return divergence. Compensation for selling protection against big overnight quarticity is smaller than in the case of overnight return divergence. The third finding is about the conditional magnitudes of premia for trading big risk. I find that in night-time trading profits from big skewness and big quarticity strategies are, respectively, significantly higher and lower, if the preceding daytime-measured jump variation (in any direction) falls in its top decile. In those strategies, the aggregate option cost is the dominant component of the final excess payoff. This, together with the fact that I do not observe clusterings of big risk events, suggests that after a "jump" investors hike their expectations of another jump occurring, which is reflected in option prices. The effect is persistent. This evidence is similar to findings of Andersen, Fusari, and Todorov, 2015b who find that

put options become – and stay – more expensive after a negative jump in returns.

Overall, my empirical investigation brings around a new method of direct measurement of compensation for big risk, often identified with jump risk. Through this lens I report novel evidence on risk premia available to investors in the market for European index options. The evidence raises new challenges for theoretical asset pricing models, as well as for the option pricing literature.

2 Arbitrage free dispersion

2.1 Introduction

Arbitrage-free markets are characterized by tight relations between unobservable pricing kernels, observable asset payoffs and their arbitrage-free price. These relations constrain the joint stochastic properties of pricing kernels and asset returns along several dimensions, which are informative about the market price of the relevant uncertain economic states and the set of arbitrage-free prices of untraded assets in incomplete markets. In this paper, we introduce a new systematic approach for testing asset pricing models, which parsimoniously aggregates the observable asset pricing information from multivariate arbitrage-free markets and comprehensively characterize tight arbitrage-free constraints on the joint distribution of potentially multivariate pricing kernels and asset returns. We complement and extend the existing literature, by introducing a new systematic approach to test multivariate pricing kernel specifications on multiple investment horizons.

We first summarize the observable arbitrage-free information by a well-defined subset of observed values on the joint cumulant generating function (CGF) of pricing kernels and asset returns. To illustrate, in the simplified setting of a single pricing kernel M and a traded return R, the joint CGF is defined by

$$\mathcal{K}_{MR}(m,r) := \log E[M^m R^r] \; ; \; (m,r) \in \mathbb{R}^2 \; , \tag{2.1.1}$$

and the asset pricing constraint for the traded return is $\mathcal{K}_{MR}(1,1) = \log E[MR] = 0$. When the marginal distribution of returns is also observable, then $\mathcal{K}_{MR}(0,\cdot)$ is observed and the observable information on $\mathcal{K}_{MR}$ is summarized by the values of the CGF on the observable set $\mathcal{O}_{\mathcal{K}_{MR}} := \{(m,r) \in dom(\mathcal{K}_{MR}) : m = 0 \text{ or } (m,r) = (1,1)\}$.

Second, we derive a broad class of multivariate arbitrage-free inequalities between observable and unobservable regions of the joint CDF, which are a consequence of the convexity of cumulant generating functions. In this way, the observable information on traded asset returns restricts the class of joint distributions for pricing kernels and returns that are consistent with arbitrage-free markets.[1] We also show that while convex arbitrage-free inequalities hold and are computable for general multivariate settings, their application to lower dimensional settings provides

[1] For instance, it constraints the marginal distribution of pricing kernels, by bounding the range of admissible nonlinear moments. Similarly, it constrains the range of admissible prices for untraded nonlinear payoffs, such as the prices of payoffs with nonlinear exposure in the underlying return.

a direct derivation of a large class of pricing kernel bounds in the literature. For instance, convexity of $\mathcal{K}_{MR}$ in equation (2.1.1) implies for any $\alpha \in (0,1)$ the bound:

$$\begin{aligned} \log E[M^\alpha] = \mathcal{K}_{MR}(\alpha, 0) &\leq \alpha \mathcal{K}_{MR}(1,1) + (1-\alpha)\mathcal{K}_{MR}(0, -\alpha/(1-\alpha)) \\ &= \log E[R^{-\alpha/(1-\alpha)}]^{1-\alpha} \,, \end{aligned} \tag{2.1.2}$$

which naturally implies $\mathcal{E}(M) := -\log E[M/E[M]] \geq \log(E[R]E[M])$, i.e., the entropy pricing kernel bound in Bansal and Lehmann, 1997, Alvarez and Jermann, 2005, Liu, 2015 and Backus, Chernov, and Zin, 2014, among others.[2]

Third, we show that convex arbitrage-free inequalities are interpretable as arbitrage-free constraints on the multivariate dispersion of pricing kernels and returns, where dispersion is defined using a family of so-called Jensen's gaps generated by the multivariate CGF. In this sense, violations of certain convex inequalities or pricing kernel bounds are equivalent to an insufficient arbitrage-free dispersion in some regions of the multivariate state space. To illustrate, given CGF (2.1.1) and a non degenerate prior distribution π, Jensen's inequality induces the Jensen's gap:

$$\mathcal{J}_\pi(M, R) := E_\pi[\mathcal{K}_{MR}(m, r)] - \mathcal{K}_{MR}(E_\pi[(m, r)]) \geq 0 \,, \tag{2.1.3}$$

with equality if and only if $(M, R) = E[(M, R)]$. Jensen's gap $\mathcal{J}_\pi(M, R)$ is a measure of multivariate dispersion and directly contrains the arbitrage-free CGF whenever $E_\pi[\mathcal{K}_{MR}(m, r)]$ or $\mathcal{K}_{MR}(E_\pi[(m, r)])$ is observable.[3] In this way, we identify unique model-free constraints for the joint distribution of pricing kernels and returns that are directly interpretable as specification constraints for their joint dispersion properties.

Fourth, we parsimoniously incorporate arbitrage-free dispersion constraints into lower and upper bounds on the CGF of pricing kernels and returns. Precisely, we introduce the upper (lower) arbitrage-free CGF $\mathcal{K}^U_{MR}$ ($\mathcal{K}^L_{MR}$), which is defined as the smallest (largest) observable upper (lower) bound implied by convex arbitrage-free inequalities on any arbitrage-free CGF, and we naturally derive upper (lower) arbitrage-free CGFs from basic arbitrage-free dispersion properties. To illustrate, given an arbitrage-free CGF evaluated at point $(m_\star, r_\star)$ and a prior with support in $\mathcal{O}_{\mathcal{K}_{MR}}$ such that $(m_\star, r_\star) = E_{\pi_\star}[(m, r)]$, inequality (2.1.3) gives:

$$E_{\pi_\star}[\mathcal{K}_{MR}(m, r)] \geq \mathcal{K}_{MR}(m_\star, r_\star) \,. \tag{2.1.4}$$

[2] The bound follows by continuity, taking limits as $\alpha \to 1$ in the convex arbitrage-free inequality (2.1.2) .

[3] For instance, the inequality $E_\pi[\mathcal{K}_{MR}(m, r)] \geq \mathcal{K}_{MR}(E_\pi[(m, r)])$ implied by a Bernoulli prior with $\pi(1,1) = \alpha \in (0,1)$ and $\pi(0, -\alpha/(1-\alpha)) = 1-\alpha$ directly generates pricing kernel bound bound (2.1.2). We discuss in detail below the properties of $\mathcal{J}_\pi(M, R)$ as a measure of multivariate dispersion.

As the left side of this inequality is observable, we obtain the following upper arbitrage-free CGF $\mathcal{K}^U_{MR}$ evaluated in $(m_\star, r_\star)$:[4]

$$\mathcal{K}^U_{MR}(m_\star, r_\star) := \inf_{\pi_\star} E_{\pi_\star}[\mathcal{K}_{MR}(m, r)] \geq \mathcal{K}_{MR}(m_\star, r_\star) \,. \tag{2.1.5}$$

Upper and lower arbitrage-free CGFs are observable in a model-free way and naturally constrain the joint distribution of pricing kernels and returns. We show in a number of important cases that these constraints are the tightest attainable, given an observed arbitrage-free price system. Thus, upper and lower CGFs meaningfully synthesize the available model-free information on the distribution of pricing kernels and returns that can be inferred from statistical return observation and existing arbitrage-free relations.

We make use of our theory of AFD to specify a coherent diagnostics approach for systematically testing the multivariate dispersion properties of asset pricing models. We apply this diagnostics approach to test the multivariate dispersion properties of transient and martingale pricing kernel components in the benchmark Bansal and Yaron, 2004-model of Long Run Risks (LRR), where the transient component is approximated empirically from the observed return of a long-maturity bond. While the marginal dispersion properties of pricing kernel components are broadly consistent with the data, multivariate dispersion tests provide sharper evidence of a model failure, in terms of (i) a counterfactural dependence of long vs. short maturity bond returns and (ii) an insufficient dispersion for pricing the optimal returns of a class of power utility investors invested in short-term bonds and aggregate market equity.

The remainder of the paper proceeds as follows. In Section 2.2, we introduce the joint CGF of pricing kernels and asset returns, showing that a number of interesting model settings – including multiple SDF components, countries and time horizons – can naturally be incorporated in this framework. Section 2.3 exploits the convexity of the joint CGF to specify general multivariate dispersion measures in a variety of relevant asset pricing contexts. Section 2.4 shows how they induce natural model-free bounds on the arbitrage-free CGF and the dispersion of pricing kernels and returns, motivating a systematic diagnostics approach for testing the multivariate dispersion implications of asset pricing models. Section 2.5 addresses in more detail concrete asset pricing contexts and explicitly computable dispersion constraints in these settings. It demonstrates the sharpeness of pricing kernel dispersion bounds resulting from our approach in a number of important cases and delivers testable multivariate dispersion constraints for the benchmark Long Run Risk (LRR) model in Bansal and Yaron, 2004. Section 2.6 characterize and tests empirically the multivariate arbitrage-free dispersion properties of recent empirical parameterizations of LRR models. Section 2.7 concludes.

[4] The infimum on the left hand side of this inequality is over priors with support in $\mathcal{O}_{\mathcal{K}_{MR}}$ such that $(m_\star, r_\star) = E_{\pi_\star}[(m, r)]$.

2.2 Arbitrage Free Cumulant Generating Function

Under weak assumptions, arbitrage-free markets imply tight relations between unobservable pricing kernels and the arbitrage-free price of marketed asset payoffs. Such constraints are conveniently summarized by particular restrictions on the joint CGF of pricing kernels and payoffs.

2.2.1 Definition

We introduce the joint CGF of pricing kernel and asset returns using a general multivariate structure, in which uncertainty is generated by a vector of returns priced by another vector of pricing kernel components. Examples of pricing kernel components are vectors of domestic and foreign pricing kernels in international arbitrage-free markets, the vector of single-period pricing kernels that price returns over different horizons or the distinct frequency components of a pricing kernel, as, e.g., in Alvarez and Jermann (2005).

Definition 3 (CGF of Pricing Kernel and Returns)**.** *Given a set of strictly positive pricing kernel components $M := (M_1, \ldots, M_{d_1})$ and a set of positive marketed gross returns $R := (R_1, \ldots, R_{d_2})$, the arbitrage-free cumulant generating function of (M, R) is the function $\mathcal{K}_{MR} : \mathbb{R}^{d_1+d_2} \to \overline{\mathbb{R}}$, defined for any $m := (m_1, \ldots, m_{d_1})$ and $r := (r_1, \ldots, r_{d_2})$ by*

$$\mathcal{K}_{MR}(m, r) := \log E\left[M^r R^r\right] := \log E\left[\prod_{i=1}^{d_1} M_i^{m_i} \prod_{j=1}^{d_2} R_j^{r_j}\right] . \tag{2.2.1}$$

The marginal pricing kernel (returns) CGF is defined by $\mathcal{K}_M(\cdot) := \mathcal{K}_{MR}(\cdot, 0)$ ($\mathcal{K}_R(\cdot) := \mathcal{K}_{MR}(0, \cdot)$).

The joint CGF uniquely identifies the joint distribution of pricing kernel components and returns.[5] Therefore, it also identifies the arbitrage-free pricing system implied by a (parametric or nonparametric) specification of an asset pricing model. Conversely, the empirically observable arbitrage-free pricing restrictions are naturally summarized by the values of an arbitrage-free CGF on a corresponding subset of points $(m, r) \in \mathbb{R}^{d_1+d_2}$. Such restrictions generate natural specification constraints for asset pricing models.

Definition 4 (Observable Arbitrage-Free CGF)**.** *An arbitrage-free CGF is observable in $(m, r) \in \mathbb{R}^{d_1+d_2}$, whenever $\mathcal{K}_{MR}(m, r)$ is known through the statistical observation of asset returns or the prices of extant payoffs. We denote by $\mathcal{O}_{\mathcal{K}_{MR}} := \{(m, r) \in \mathbb{R}^{d_1+d_2} : \mathcal{K}_{MR}(m, r)$ is observed$\}$ the set of observable points of an arbitrage-free CGF and call $\mathcal{K}_{MR}|_{\mathcal{O}_{\mathcal{K}_{MR}}}$ the observable aribtrage-free CGF.*

[5] Throughout the paper we assume that the joint cumulant generating function is finite in a non-degenerate open domain containing the origin.

2.2.2 Observability and Marginal CGF

Whenever we can assume statistical observability of the return distribution, the marginal return CGF is observable, i.e., $(0, r) \in \mathcal{O}_{\mathcal{K}_{MR}}$ for any $r \in \mathbb{R}^{d_2}$. In general, the marginal CGF of a pricing kernel is never empirically completely observable. Whenever the price B of a risk-free zero bond is observable, then $\mathcal{K}_M(1) = \log E[M] = \log B$ and $(1, 0) \in \mathcal{O}_{\mathcal{K}_{MR}}$. Additional points on the marginal CGF may be directly observable due to normalizing conditions. For instance, when a pricing kernel is decomposed into the product of a transient and a permanent martingale component M^T and M^P, the martingale condition $\mathcal{K}_{M^T M^P}(0, 1) = \log E[M^P] = 0$ yields $(0, 1, 0) \in \mathcal{O}_{\mathcal{K}_{M^T M^P R}}$; see again Alvarez and Jermann (2005), among others.

2.2.3 Univariate Return and Pricing Kernel ($d_1 = d_2 = 1$)

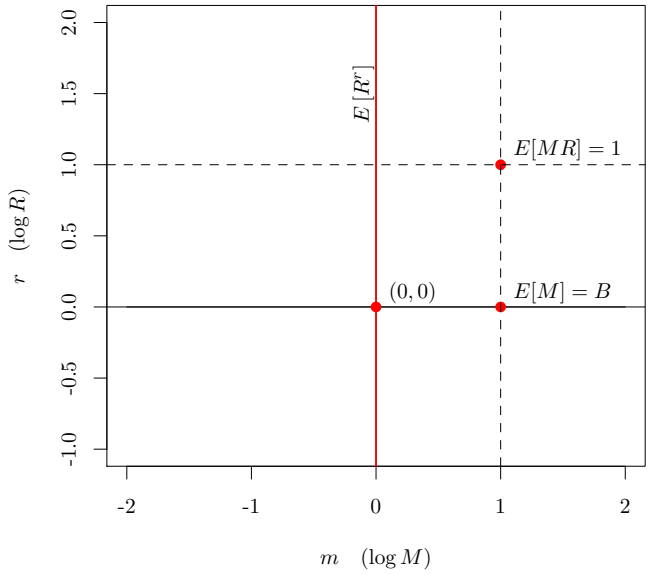
Given the statistical observability of a univariate return distribution, the additional set of observable points depends on the structure of observable arbitrage-free constrains. Whenever a risk-free bond is priced, then obviously $(1, 0) \in \mathcal{O}_{\mathcal{K}_{MR}}$. Whenever a risky return is also priced, then $\mathcal{K}_{MR}(1, 1) = \log E[MR] = 0$ and $(1, 1) \in \mathcal{O}_{\mathcal{K}_{MR}}$. Figure 2.1a plots the observable set $\mathcal{O}_{\mathcal{K}_{MR}}$ generated by statistical return information and such arbitrage-free constraints. This set is not convex. It contains the vertical line with abscissa in $m = 0$, i.e., the domain of the observable moment generating function of returns, and only two additional points of the vertical line with abscissa in $m = 1$, which is the region collecting information about the risk-neutral distribution of returns. The sparsity of points in this region reflects the large degree of market incompleteness of this setting, in which no nonlinear payoff on underlying return R is traded. The other extreme is a market extended to complete option trading, allowing to trade any smooth nonlinear function of R using portfolios of out-of-the-money options. In such a case, we have for any $p \in \mathbb{R}$, following Carr and Madan, 1998 and Schneider and Trojani, 2015a:

$$\mathcal{K}_{MR}(1, p) = \log E[MR^p] = \log[p + (1 - p)B + p(p - 1) \int_0^\infty K^{p-2} O(R, K) dK] \,,$$

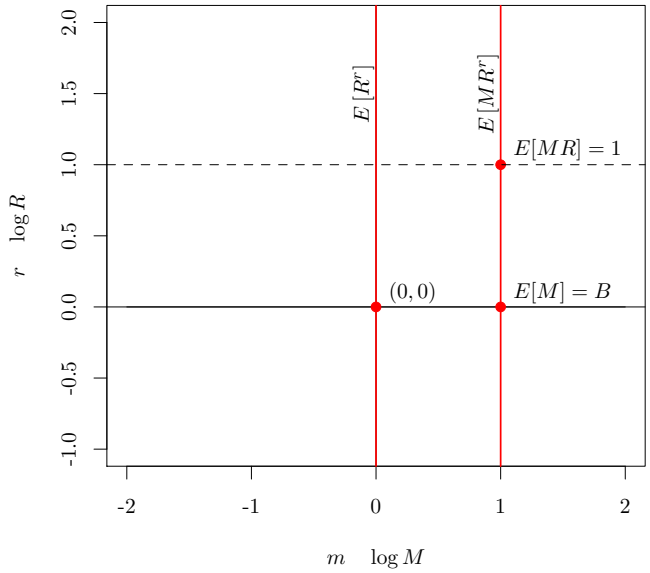
where $O(R, K)$ is the arbitrage-free price of an out-of-the-money option on R with strike price $K > 0$. Figure 2.1b illustrates the enrichment of the (non convex) set of empirically observable CGF points generated by complete option markets, which now includes the vertical line with abscissa in $m = 1$.

2.2.4 Transient and Persistent Pricing Kernel Components ($d_1 = 2$ and $d_2 = 1$)

Whenever the long end of the yield curve is observable, the pricing kernel decomposition $M^P M^T$ into transient and persistent components produces additional observable constraints. Following Alvarez and Jermann (2005), M^P does not affect the price of infinitely long maturity zero coupon bonds and $R_\infty = 1/M^T$, where R_∞ is



(A) $\mathcal{O}_{\mathcal{K}_{MR}}$ with incomplete option markets.



(B) $\mathcal{O}_{\mathcal{K}_{MR}}$ with complete option markets.

FIGURE 2.1: Observable sets in (M, R) space for $d_1 = d_2 = 1$. The red points and segments represent the tuples $(m, r) \in \mathcal{O}_{\mathcal{K}_{MR}} \subset \mathbb{R}^2$ where the joint arbitrage-free CGF is observable, based on statistical return observations and asset pricing restrictions on the risk-free bond and the risky asset (panel (a)). In panel (b) the additional observable points on the $m = 1$ line result from observing the prices of a continuum of options which can be used to replicate portfolios with final payoff R^r.

the return of the infinitely long maturity zero coupon bond. Therefore,

$$\mathcal{K}_{MR}(m_T, 0, r) = \log E[(M^T)^{m_T} R^r] = \log E[R_\infty^{-m_T} R^r] \,, \tag{2.2.2}$$

where $M := (M^T, M^P)$ and $(m_T, 0, r) \in \mathcal{O}_{\mathcal{K}_{MR}}$ for each $(m_T, r) \in \mathbb{R}^2$. The pricing constraints for the short term zero bond and the risky return imply $(1, 1, 0), (1, 1, 1) \in$

$\mathcal{O}_{\mathcal{K}_{MR}}$. Non convex set $\mathcal{O}_{\mathcal{K}_{MR}}$ is illustrated in Figure 2.2. The observable points resulting from arbitrage-free pricing relations are marked as violet circles (○), while those that are observable statistically span the red (m_T, r) plane. Two such points, corresponding to the CGF values $\mathcal{K}_{MR}(0,0,1) = \log E[R]$ and $\mathcal{K}_{MR}(1,0,0) = \log E\left[R_\infty^{-1}\right]$, are marked as violet squares (□).

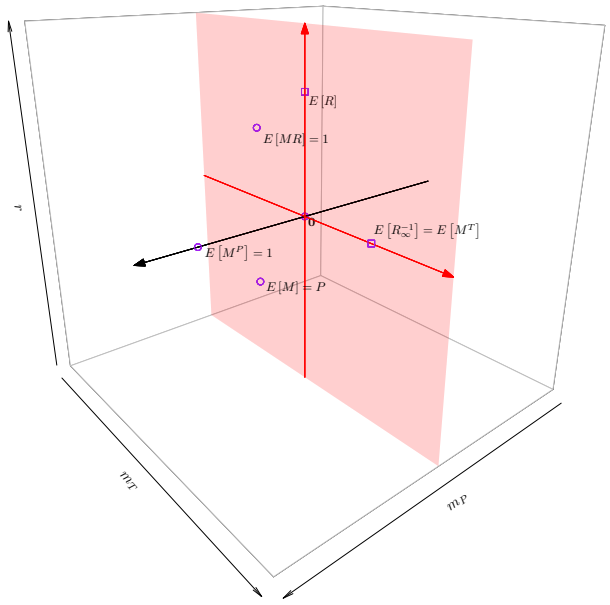


FIGURE 2.2: Observable points $\mathcal{O}_{\mathcal{K}_{MR}}$ of the joint CGF in the single risky asset case with persistent and transient pricing kernel components. The red surface and the purple points represent the tuples $(m, r) \in \mathcal{O}_{\mathcal{K}_{MR}} \subset \mathbb{R}^3$ where the joint CGF is observed based on statistical observations and asset pricing restrictions on the short-term risk-free bond, the long-term risk-free bond and the risky asset.

2.2.5 Domestic and Foreign Pricing Kernels ($d_1 = 2$ and $d_2 = 2$)

In an international context, the pricing kernel components can be domestic and foreign pricing kernels M_d and M_f, pricing domestic and foreign risk-free bonds and risky returns. The empirically observable points follow as in the previous examples, so that points $(1,0,0,0)$, $(0,1,0,0)$, $(1,0,1,0)$ and $(0,1,0,1)$ are all elements of $\mathcal{O}_{\mathcal{K}_{MR}}$, where $M := (M_d, M_f)$ and $R := (R_d, R_f)$ is the vector of domestic and foreign risky returns. Whenever domestic and foreign option markets are complete, it also follows $(1,0,r_d,0), (0,1,0,r_f) \in \mathcal{O}_{\mathcal{K}_{MR}}$ for any $(r_d, r_f) \in \mathbb{R}^2$. Additional observable constraints can emerge from the spot exchange rate market, as the exchange rate return $R_e := (M_f/M_d) \cdot \mathcal{E}$ needs to satisfy the additional arbitrage-free conditions:

$$1 = E[M_d R_e] = E[M_f \mathcal{E}] \; ; \; 1 = E[M_f (1/R_e)] = E[M_d (1/\mathcal{E})] \; .$$

When domestic and foreign markets are complete, $R_e = M_d/M_f$ and the joint CGF of pricing kernels and returns characterizes the observable arbitrage-free restrictions from domestic, foreign and spot exchange rate markets. Further CGF constraints

arise when complete exchange rate option markets allow the trading of smooth functions of R_e. Indeed, in this case $(1-p, p, 0, 0) \in \mathcal{O}_{\mathcal{K}_{MR}}$ for every $p \in \mathbb{R}$, since

$$\mathcal{K}_{MR}(1-p, p, 0, 0) = \log E[M_d R_e^p] = \log[p + (1-p)B_d + p(p-1)\int_0^\infty K^{p-2} O(R_e, K) dK] .$$

A non degenerate exchange rate component $\mathcal{E}$ due to market incompleteness can also be naturally incorporated into our framework, by means of an arbitrage-free joint CGF of variables $(M_d, M_f, R_d, R_f, \mathcal{E})$ and the corresponding non convex set of observable points.

2.2.6 Multi-Period Pricing Kernels and Returns ($d_1 = d_2$)

Given an investment horizon consisting of $d = d_1 = d_2$ periods, we can easily incorporate multi-period arbitrage-free information into our framework. Let $\{M_i : i = 1, \ldots, d\}$ ($\{R_i : i = 1, \ldots, d\}$) be a sequence of single-period pricing kernels (risky returns) for pricing time i payoffs at time $i-1$ (priced at time $i-1$ and paying off at time i). Whenever risk-free bond prices B_i for maturity $i = 1, \ldots, d$ are observed, then arbitrage-free CGF yields:

$$\mathcal{K}_{MR}(\iota_i, 0_{2d-i}) = \log E\left(\prod_{k=1}^{i} M_k\right) = \log B_i , \tag{2.2.3}$$

where ι_i is a vector of ones in $\mathbb{R}^i$ and 0_{2d-i} a vector of zeros in $\mathbb{R}^{2d-i}$, i.e., $(\iota_i, 0_{2d-i}) \in \mathcal{O}_{\mathcal{K}_{MR}}$. Similarly,

$$\mathcal{K}_{MR}(\iota_i, 0_{d-i}, \iota_i, 0_{d-i}) = \log E\left[\prod_{k=1}^{i} M_k R_k\right] = 0 , \tag{2.2.4}$$

i.e., $(\iota_i, 0_{d-i}, \iota_i, 0_{d-i}) \in \mathcal{O}_{\mathcal{K}_{MR}}$.

2.3 Dispersion Measured by Jensen's Gap

The convexity properties of arbitrage free CGF's are directly linked to the dispersion of pricing kernel and returns. Therefore, asset pricing restrictions are naturally formulated using appropriate measures of dispersion. In this way, we obtain a general unifying approach for testing asset pricing models.[6]

2.3.1 Jensen's Gap and Multivariate Dispersion

We propose to measure multivariate dispersion by a family of Jensen gaps implied by the joint CGF.

[6]Given that we measure dispersion by the convexity of a CGF, dispersion constraints implied by our approach are particularly appropriate to test log-linear, or nearly log-linear, models. Similarly, dispersion measured by the convexity of a CGF extends from Gaussian to non Gaussian settings in a transparent way.

Definition and Main Properties

Definition 5. *(i) For a prior distribution π on $\mathbb{R}^{d_1+d_2}$ and a joint CGF $\mathcal{K}_{MR}$, the Jensen's gap under prior π is:*

$$\mathcal{J}_\pi(M,R) := E_\pi[\mathcal{K}_{MR}(m,r)] - \mathcal{K}_{MR}(E_\pi[(m,r)]) \geq 0\,. \tag{2.3.1}$$

(ii) The marginal Jensen's gaps implied by prior π are defined similarly: $\mathcal{J}_\pi(M) := E_\pi[\mathcal{K}_M(m)] - \mathcal{K}_M(E_\pi[m])$ and $\mathcal{J}_\pi(R) := E_\pi[\mathcal{K}_R(r)] - \mathcal{K}_R(E_\pi[r])$.

$\mathcal{J}_\pi(M,R)$ ($\mathcal{J}_\pi(M)$ and $\mathcal{J}_\pi(R)$) measures (measure) the multivariate (marginal) dispersion of pricing kernels and returns, consistently with a number of useful properties, collected in Proposition 1. Among these, additivity under independent experiments is a useful property, e.g., when studying a term structure of dispersion in pricing kernels and returns.

Proposition 1. *Jensen's gap $\mathcal{J}_\pi(M,R) \geq 0$ is a dispersion measure with the following properties:*

1. *$\mathcal{J}_\pi(M,R) = 0$, whenever (M,R) is a degenerate random vector.*
2. *For any prior π with support not included in a strict subspace of $\mathbb{R}^{d_1+d_2}$, $\mathcal{J}_\pi(M,R) = 0$ if and only if (M,R) is a degenerate random vector.*
3. *Given stochastically independent pricing kernel components and returns, it follows:*

$$\mathcal{J}_\pi(M,R) = \mathcal{J}_\pi(M) + \mathcal{J}_\pi(R)\,. \tag{2.3.2}$$

4. *Given two independent random vectors (M,R) and (N,Q) in $\mathbb{R}^{d_1+d_2}$, it follows:*

$$\mathcal{J}_\pi(M\times N, R\times Q) = \mathcal{J}_\pi(M,R) + \mathcal{J}_\pi(N,Q)\,, \tag{2.3.3}$$

   where $M\times N := (M_1N_1,\ldots,M_{d_1}N_{d_1})$ and $R\times Q := (R_1Q_1,\ldots,N_{d_2}Q_{d_2})$.

5. *$\mathcal{J}_\pi(M,R)$ is positively homogenous of degree 0, in the sense that for any $\lambda_M := (\lambda_{M1},\ldots,\lambda_{Md_1})$ and $\lambda_R := (\lambda_{M1},\ldots,\lambda_{Md_2})$: $\mathcal{J}_\pi(\lambda_M\circ M, \lambda_R\circ R) = \mathcal{J}_\pi(M,R)$.*[7]
6. *For any prior π on $\mathbb{R}^{d_1+d_2}$, having prior covariance matrix $Var_\pi(m,r)$:*

$$\mathcal{J}_\pi(M,R) = \frac{tr(\Sigma Var_\pi(m,r))}{2}\,, \tag{2.3.4}$$

   whenever $(\log M, \log R)$ is a vector of Gassian variables with covariance matrix Σ.

Property 1. implies a necessary requirement for a measure of multivariate dispersion, i.e., that it is zero whenever (M,R) is a degenerate random vector. This property follows from the linearity of the joint CGF in such a case. From Property

[7] $\circ$ denotes the Hadamard product. For two $n\times m$ matrices A,B with elements A_{ij}, B_{ij}, matrix $A\circ B$ is the $n\times m$ matrix with elements $(A\circ B)_{ij} = A_{ij}B_{ij}$.

2, a strictly positive Jensen's gap is generated by the strict convexity of the joint CGF of multivariate non degenerate random variables, whenever prior π is not concentrated on a strict subspace of $\mathbb{R}^{d_1+d_2}$. In this case, $\mathcal{J}_\pi(M,R) = 0$ if and only if random vector (M,R) is degenerate. Property 3. is an additive decomposition of the joint Jensen's gap into the sum of marginal Jensen's gaps, when pricing kernel and returns are stochastically independent. Property 4. implies additivity of $\mathcal{J}_\pi(M,R)$ under independent experiments, a desirable aggregation property that generalizes the additivity of univariate dispersion measures such as entropy; see, e.g., Backus, Chernov, and Zin, 2014. Property 5. implies scale invariance, while Property 6. gives the expression for $\mathcal{J}_\pi(M,R)$ in the benchmark case of jointly Gaussian log pricing kernel and returns.

Jensen's Gap Dimension

The support of prior π in equation (2.3.1) introduces a degree of flexibility in $\mathcal{J}_\pi(M,R)$, which can be used to localize dispersion on particular regions of a multivariate subspace. An obvious localization is on the marginals of (M,R). More generally, a prior with support in a subspace of dimension $d < d_1 + d_2$ can measures the dispersion of particular linear combinations of pricing kernel and returns.

Definition 6. *(i) Given a prior distribution with prior covariance matrix $Var_\pi(m,r)$ such that $0 < tr(Var_\pi(m,r)) < \infty$, we call $d_\pi := rank(Var_\pi(m,r))$ the dimension of $\mathcal{J}_\pi(M,R)$. (ii) Given a Jensen's gap of dimension d_π, a standardized Jensen's gap of dimension d_π is defined by*

$$\mathcal{D}_\pi(M,R) := \frac{\mathcal{J}_\pi(M,R)}{tr(Var_\pi(m,r))} \,. \tag{2.3.5}$$

When $d_\pi < d_1 + d_2$ the prior π is concentrated on a $d_\pi-$dimensional subspace of $\mathbb{R}^{d_1+d_2}$. As a consequence, if $\mathcal{J}_\pi(M,R) = 0$ the components of random vector $(\log M, \log R)$ are related by an affine deterministic relationship. According to Proposition 1, a standardized Jensen's gap satisfies the convenient normalization $\mathcal{D}_\pi(M,R) = \sigma^2/2$, whenever $(\log M, \log R)$ is an iid vector of Gaussian variables with variance σ^2. In presence of deviations from Gaussianity, the leading contribution to $\mathcal{D}_\pi(M,R)$ is approximatively given by

$$\mathcal{D}_\pi(M,R) \approx \frac{tr(\mathcal{K}''_{MR}(E_\pi[m], E_\pi[r]) Var_\pi(m,r))}{2tr(Var_\pi(m,r))} \,, \tag{2.3.6}$$

where $\mathcal{K}''_{MR}(\cdot,\cdot)$ is the Hessian of $\mathcal{K}$.[8] This approximation is exact for Gaussian random vectors and likely sufficiently accurate for priors with moderate degree of multivariate skewness and kurtosis.

2.3.2 Jensen's Gap and Entropy Measures

Jensen's gaps extend well-known concepts of dispersion in the literature, such as several useful measures of entropy and co-entropy.[9] They also induce a broad family of new concrete measures, such as, e.g., generalized entropy and generalized co-entropy, dispersion measures linked to Chernoff, 1952 information, or asymmetric measures of co-dispersion. We illustrate their properties in the context of simple asset pricing settings.

Univariate Return and Pricing Kernel ($d_1 = d_2 = 1$)

Generalized Entropy. Given a Bernoulli prior π_α with mass $\alpha \in (0,1)$ in $(m,r) = (1,0)$ and mass $1-\alpha$ in $(m,r) = (0,0)$, we obtain:

$$\begin{aligned} \mathcal{D}_{\pi_\alpha}(M,R) &= \frac{\alpha\mathcal{K}_{MR}(1,0) + (1-\alpha)\mathcal{K}_{MR}(0,0) - \mathcal{K}_{MR}(\alpha,0)}{\alpha(1-\alpha)} \\ &= \frac{\log E[(M/E(M))^\alpha]}{\alpha(\alpha-1)} =: \mathcal{E}_\alpha(M)\,, \end{aligned} \tag{2.3.8}$$

i.e., the $\alpha-$Rényi, 1960 entropy of the stochastic discount factor. $\mathcal{E}_\alpha(M)$ is a standardized Jensen's gap of dimension $d_\pi = 1$. It's geometric interpretation based on dispersion measure $\mathcal{D}_{\pi_\alpha}(M,R)$ is illustrated in Figure 2.3.

As $\alpha \to 0$, we obtain $\mathcal{E}_0(M) = E[-\log(M/E(M))]$, i.e., the stochastic discount factor entropy in Alvarez and Jermann, 2005, Backus, Chernov, and Martin, 2011, Backus, Chernov, and Zin, 2014, among others.

Co-Generalized Entropy. When returns and pricing kernel are stochastically independent, Proposition 1 implies $\mathcal{E}_\alpha(MR) = \mathcal{E}_\alpha(M) + \mathcal{E}_\alpha(R) =: \mathcal{E}^\perp_\alpha(MR)$. We can define the $\alpha-$Rényi co-entropy by

$$\mathcal{C}_\alpha(M,R) := \frac{\alpha-1}{\alpha}\Big(\mathcal{E}_\alpha(MR) - \mathcal{E}^\perp_\alpha(MR)\Big) = \frac{1}{\alpha^2}\log\left(\frac{E[(MR)^\alpha]}{E[M^\alpha]E[R^\alpha]}\right), \tag{2.3.9}$$

[8] Higher order contributions can be computed following the multivariate cumulant expansion approach in Jammalamadaka, Rao, and Terdik, 2006. Note that

$$\mathcal{D}_\pi(M,R) \approx \sum_{i=1}^{d_\pi} \frac{\lambda_{i\pi}}{tr(Var_\pi(m,r))} \cdot \langle q_{i\pi}, \mathcal{K}''_{MR}(E_\pi[m], E_\pi[r]) q_{i\pi}\rangle\,, \tag{2.3.7}$$

with the nonzero eigenvalues $\lambda_{i\pi}$ and the corresponding orthonormal eigenvectors $q_{i\pi}$ of $Var_\pi(m,r)$, denoting by $\langle\cdot,\cdot\rangle$ the Euclidean scalar product. It follows that whenever dimension d_π is strictly less than $d_1 + d_2$, some linear combinations of the columns of $\mathcal{K}''_{MR}(E_\pi[m], E_\pi[r])$ do not contribute to the right hand of approximation (2.3.7). For instance, in the one dimensional case, $\mathcal{D}_\pi(M,R)$ can be approximated by a single quadratic form of matrix $\mathcal{K}''_{MR}(E_\pi[m], E_\pi[r])$.

[9] See, e.g., Alvarez and Jermann, 2005, Backus, Chernov, and Martin, 2011, Backus, Chernov, and Zin, 2014, Backus, Boyarchenko, and Chernov, 2016, Bakshi and Chabi-Yo, 2014 and Chabi-Yo and Colacito, 2013, among others.

which for $\alpha \to 0$ yields the co-entropy introduced in Backus, Boyarchenko, and Chernov, 2016 and Bakshi and Chabi-Yo, 2014, among others. Note that whenever $(\log M, \log R)$ is jointly normally distributed, Proposition 1 implies $\mathcal{C}_\alpha(M,R) = Cov(\log M, \log R)$.

Power Generalized Entropy. Given a Bernoulli prior π_α with mass $\alpha \in (0,1)$ in $(m,r) = (p,0)$ ($p \in \mathbb{R}$) and mass $1-\alpha$ in $(m,r) = (0,0)$, we obtain

$$\mathcal{D}_{\pi_\alpha}(M,R) = \frac{\log E[(M^p/E(M^p))^\alpha]}{p^2\alpha(\alpha-1)} = \frac{1}{p^2}\mathcal{E}_\alpha(M^p) =: \mathcal{E}^p_\alpha(M)\,, \tag{2.3.10}$$

i.e., $\mathcal{D}_{\pi_\alpha}(M,R)$ is proportional to the $\alpha-$Rényi, 1960 entropy of the $p-$th power of the pricing kernel. $\mathcal{E}^p_\alpha(M)$ is a standardized Jensen's gap of dimension $d_\pi = 1$. For $\alpha \to 0$ we obtain $4\mathcal{E}^2_0(M) = \mathcal{E}_0(M^2)$, i.e., the entropy of the squared pricing kernel adopted in Bakshi and Chabi-Yo, 2014 to specify tractable multivariate pricing kernel bounds.[10]

Dimension $d_\pi > 1$. Most dispersion measures rely on a dimension $d_\pi = 1$. Jensen's gaps of dimension $d_\pi > 1$ are easily constructed. To illustrate, consider a prior $\pi_{\alpha,\beta}$ with mass $\alpha > 0$ in $(m,r) = (1,0)$, mass $\beta > 0$ in $(m,r) = (0,1)$ and mass $1-(\alpha+\beta)$ in $(m,r) = (0,0)$. It then follows:

$$\begin{aligned}\mathcal{D}_{\pi_{\alpha,\beta}}(M,R) &= \frac{\alpha\mathcal{K}_{MR}(1,0) + \beta\mathcal{K}_{MR}(0,1) - \mathcal{K}_{MR}(\alpha,\beta)}{\alpha(1-\alpha)+\beta(1-\beta)} \\ &= \frac{\log E[(M/E(M))^\alpha (R/E[R])^\beta]}{\alpha(\alpha-1)+\beta(\beta-1)} =: \mathcal{E}_{\alpha,\beta}(M,R)\,. \end{aligned} \tag{2.3.11}$$

$\mathcal{E}_{\alpha,\beta}(M,R)$ is a standardized Jensen's gap of dimension $d_\pi = 2$ and a proper measure of bivariate dispersion. Independence of pricing kernel and returns additionally implies

$$\mathcal{E}_{\alpha,\beta}(M,R) = \frac{\alpha(\alpha-1)\mathcal{E}_\alpha(M) + \beta(\beta-1)\mathcal{E}_\beta(R)}{\alpha(\alpha-1)+\beta(\beta-1)} =: \mathcal{E}^\perp_{\alpha,\beta}(M,R)\,, \tag{2.3.12}$$

i.e., $\mathcal{E}^\perp_{\alpha,\beta}(M,R)$ equals a convex combination of Rényi, 1960 entropies. A measure of co-dispersion that (i) is zero if and only if M and R are stochastically independent and (ii) equals $Cov(\log M, \log R)$ when returns and pricing kernel are jointly log normal, then naturally follows:

$$\mathcal{C}_{\alpha,\beta}(M,R) := \frac{\alpha(\alpha-1)+\beta(\beta-1)}{\alpha\beta}(\mathcal{E}_{\alpha,\beta}(M,R) - \mathcal{E}^\perp_{\alpha,\beta}(M,R)) = \frac{1}{\alpha\beta}\log\left(\frac{E\left[M^\alpha R^\beta\right]}{E\left[M^\alpha\right]E\left[R^\beta\right]}\right)\,.$$

$\mathcal{C}_{\alpha,\beta}(M,R)$ is in general not a symmetric measure of co-dispersion. This property can be useful, e.g., to characterize pricing kernel and return dependence while explicitly

[10] Using power generalized entropy and Proposition 1, it is also possibe to specify convenient measures of co-dispersion, which are consistent with Pearson's measure of correlation in the log Gaussian case.

accounting for the asymmetric role of pricing kernels and individual asset returns in arbitrage-free markets.[11]

Domestic and Foreign Pricing Kernels ($d_1 = 2$ and $d_2 = 2$)

Chernoff, 1952 Information. Given a Bernoulli prior π_α with mass $\alpha \in (0,1)$ in $(m,r) = (1,0,0,0)$ and mass $1-\alpha$ in $(0,1,0,0)$, we obtain:

$$\begin{aligned}\mathcal{J}_{\pi_\alpha}(M,R) &= \alpha\mathcal{K}_{MR}(1,0,0,0) + (1-\alpha)\mathcal{K}_{MR}(0,1,0,0) - \mathcal{K}_{MR}(\alpha,1-\alpha,0,0) \\ &= -\log E[(M_d/E(M_d))^\alpha (M_f/E(M_f))^{1-\alpha}] =: -\log CC_\alpha(M_d,M_d)\ ,\end{aligned}$$

with the $\alpha-$Chernoff, 1952 coefficient $CC_\alpha(M_d,M_d)$. The optimal Chernoff, 1952 coefficient

$$\mathcal{CC}_{\alpha^*}(M_d,M_f) := \min_{\alpha\in(0,1)} \mathcal{CC}_\alpha(M_d,M_f)\ , \tag{2.3.13}$$

is a symmetric measure of similarity between pricing kernels, while Chernoff, 1952 information (or Chernoff divergence):

$$\mathcal{CI}_*(M_d,M_f) := -\ln \mathcal{CC}_{\alpha^*}(M_d,M_f) = \max_{\alpha\in(0,1)} \mathcal{J}_{\pi_\alpha}(M,R)\ , \tag{2.3.14}$$

is a symmetric measure of discrepancy between pricing kernels.

2.4 Informative and Observable Arbitrage Free Dispersion

The convexity of the joint CGF imposes constraints on the dispersion properties of pricing kernels and returns. Therefore, we make use of Jensen's gaps to characterize the testable dispersion properties that any asset pricing model needs to satisfy.

2.4.1 Definition

If quantities $E_\pi[\mathcal{K}_{MR}(m,r)]$ and $\mathcal{K}_{MR}(E_\pi[(m,r)])$ in Definition 5 are directly computable from the known values of $\mathcal{K}_{MR}$ on observable set $\mathcal{O}_{\mathcal{K}_{MR}}$, then $\mathcal{J}_\pi(M,R)$ is observable and directly produces verifiable constraints on the convexity features of any arbitrage-free CGF. The situation is different when either $E_\pi[\mathcal{K}_{MR}(m,r)]$ or $\mathcal{K}_{MR}(E_\pi[(m,r)])$ is unobservable.

Definition 7. *(i) Jensen's gap $\mathcal{J}_\pi(M,R)$ in Definition 5 is an informative arbitrage-free dispersion whenever (1) π has support in $\mathcal{O}_{\mathcal{K}_{MR}}$ or (2) $E_\pi[(m,r)] \in \mathcal{O}_{\mathcal{K}_{MR}}$. (ii) Inequalities*

$$\{\mathcal{J}_\pi(M,R) \geq 0\ :\ \textit{prior } \pi \textit{ is such that (1) holds}\}\ , \tag{2.4.1}$$

[11] By construction, $\mathcal{C}_{\alpha,\alpha}(M,R) = \mathcal{C}_\alpha(M,R)$, illustrating that lower-dimensional dispersion or co-dispersion measures are special cases of higher-dimensional dispersion or co-dispersion measures.

define the set of observable arbitrage-free dispersion constraints of Type (1). (iii) Inequalities

$$\{\mathcal{J}_\pi(M,R) \geq 0 \ : \ \textit{prior } \pi \textit{ is such that (2) holds}\}\ , \tag{2.4.2}$$

define the set of observable dispersion constraints of Type (2). (iv) An informative arbitrage-free dispersion $\mathcal{J}_\pi(M,R)$ is observable, whenever both π has support in $\mathcal{O}_{\mathcal{K}_{MR}}$ and $E_\pi[(m,r)] \in \mathcal{O}_{\mathcal{K}_{MR}}$.

Informative dispersion restricts the range of possible values of an arbitrage-free CGF in unobservable parts of its domain.[12] Observable dispersion uniquely constrains the convexity properties of the observable arbitrage-free CGF. Further observable constraints on the convexity of the arbitrage-free CGF can be obtained using observable differences of informative dispersions.

Definition 8. *Given priors π_1 and π_2 with support in $\mathcal{O}_{\mathcal{K}_{MR}}$ and such that $E_{\pi_1}[(m,r)] = E_{\pi_2}[(m,r)]$, $\Delta\mathcal{J}_{\pi_1,\pi_2}(M,R) := \mathcal{J}_{\pi_1}(M,R) - \mathcal{J}_{\pi_2}(M,R)$ is called an observable arbitrage-free excess dispersion.*

2.4.2 Implications of Observable Dispersion and Excess Dispersion

Observable dispersions and excess dispersions naturally reflect the observability of the arbitrage-free CGF in particular regions of its domain. In order to avoid obvious model mispecifications, the model-implied and the arbitrage-free CGF have to coincide on the observable set. Therefore, the model-implied CGF convexity has to be consistent with the observable dispersion and excess dispersion. In all other cases, a dispersion violation is directly observed and a new model specification is necessary.

Definition 9. *Given model $\mathbb{M}$, an observable dispersion (excess dispersion) violation arises whenever $\mathcal{J}_\pi(M,R) \neq \mathcal{J}^{\mathbb{M}}_\pi(M,R)$ ($\Delta\mathcal{J}_{\pi_1,\pi_2}(M,R) \neq \Delta\mathcal{J}^{\mathbb{M}}_{\pi_1,\pi_2}(M,R)$) for some observable arbitrage-free dispersion $\mathcal{J}_\pi(M,R)$ (excess dispersion $\Delta\mathcal{J}_{\pi_1,\pi_2}(M,R)$).*

Transient vs. Persistent Pricing Kernel Components and Horizon Dependence

Despite the fact that observable dispersion and excess dispersion depend on directly observable convexity properties of the arbitrage-free CGF, under plausible assumptions they already characterize important properties of asset prices, including the joint dependence of permanent and transient pricing kernel components or the horizon depedence of zero-coupon bond prices.

Example 2 (Dependence of transient and persistent pricing kernel components)**.** Following Bakshi and Chabi-Yo, 2014, the covariance between transient and persistent pricing kernel components is observable when the return R_∞ of the infinite maturity

[12] This is in contrast to the observable parts of the domain, where the arbitrage-free CGF is fully identified.

bond is observable: $\text{cov}(M^T, M^P) = B - E[1/R_\infty]$. A suitable monotonic transformation and an appropriate scaling of this equality, gives:

$$\frac{1}{2}\log\left(\text{cov}(M^T/E(M^T), M^P) + 1\right) = E_{\pi_1}[\mathcal{K}_{M^T M^P}(m^T, m^P)] - E_{\pi_2}[\mathcal{K}_{M^T M^P}(m^T, m^P)]\,, \tag{2.4.3}$$

where π_1 (π_2) has mass $1/2$ in $(m^T, m^P) = (1,1)$ ($(m^T, m^P) = (1,0)$) and mass $1/2$ in $(m^T, m^P) = (0,0)$ ($(m^T, m^P) = (0,1)$). Since $E_{\pi_1}[(m^T, m^P)] = E_{\pi_2}[(m^T, m^P)] = (1/2, 1/2)$, equation (2.4.3) induces an observable excess dispersion $\Delta\mathcal{J}_{\pi_1,\pi_2}(M^T, M^P)$. In contrast to $\text{cov}(M^T, M^P)$, excess dispersion (2.4.3) is independent of the scale of M^T, i.e., its first moment. In this sense, it implies a definition of an excess dispersion violation that is robust with respect to an incorrect measurement of the first moment of $1/R_\infty$.

Example 3 (Horizon dependence)**.** Given a vector $M = (M_1, \ldots, M_n)$ of strictly stationary single-period stochastic discount factors M_i, pricing at time $i-1$ payoffs paid at time i, Backus, Chernov, and Zin, 2014 measure horizon dependence as

$$H(n) := \frac{1}{n}\mathcal{E}_0\left(\prod_{i=1}^{n} M_i\right) - \mathcal{E}_0\left(M_1\right) = \frac{\ln(B_n)}{n} - \ln(B_1)\,, \tag{2.4.4}$$

where B_i ($i = 1, \ldots, n$) is the price of a zero bond with maturity i. Therefore, $H(n)$ is a measure of the (negative) slope of the yield curve at horizon n. Denoting by ι (e_i) the vector of ones (the $i-$th unit vector) in $\mathbb{R}^n$, we can write $H(n)$ as an arbitrage-free excess dispersion:[13]

$$H(n) = \frac{\mathcal{K}_M(\iota)}{n} - \frac{\sum_{i=1}^{n}\mathcal{K}_M(e_i)}{n} = \Delta\mathcal{J}_{\pi_1,\pi_2}(M)\,, \tag{2.4.5}$$

using a prior π_1 (π_2) with mass $1/n$ in $m = \iota$ and mass $1 - 1/n$ in $m = 0_n$ (with uniform mass $1/n$ in each unit vector e_i). Thus, horizon dependence can be understood as a particular convexity requirement, along the main diagonal in the domain of the arbitrage-free CGF of M.

Implications for Observable Model-Implied CGFs

Intuitively, the absence of observable dispersion or excess dispersion violations must constrain the convexity of the arbitrage-free CGF on observable set $\mathcal{O}_{\mathcal{K}_{MR}}$ quite strongly. Indeed, in that case the observable model-implied CGF $\mathcal{K}^{\mathbb{M}}_{MR}|_{\mathcal{O}_{\mathcal{K}_{MR}}}$ is already uniquely identified, up to a linear transformation, as is stated precisely in the next proposition, proven in the Supplemental Appendix.

Proposition 2. *If there are no observable arbitrage free dispersion or excess dispersion violations, then $\mathcal{K}_{MR}|_{\mathcal{O}_{\mathcal{K}_{MR}}} - \mathcal{K}^{\mathbb{M}}_{MR}|_{\mathcal{O}_{\mathcal{K}_{MR}}}$ is linear, i.e., there exists a vector $e \in \mathbb{R}^{d_1+d_2}$ such that $\mathcal{K}_{MR}(o) = \mathcal{K}^{\mathbb{M}}_{MR}(o) + e' \cdot o$ for every $o = (m, r) \in \mathcal{O}_{\mathcal{K}_{MR}}$.*

[13] Note that $E_{\pi_1}[m] = E_{\pi_2}[m] = \iota/n$.

From Proposition 2, the absence of observable dispersion or excess dispersion violations implies an observable model-implied CGF that is uniquely identified, up to a possibly inappropriate scaling of pricing kernel components or returns. Inappropriate scaling can be corrected by rescaling, e.g., when the scaling discrepancy is concentrated in the marginal distribution of pricing kernel components. In contrast, rescaling does not correct observable dispersion or excess dispersion violation, as Jensen's gap is homogenous of degree zero (see point 5. of Proposition 1). This motivates the next definition.

Definition 10 (Scaling Discrepancy)**.** *Whenever $\mathcal{K}_{MR}|_{\mathcal{O}_{\mathcal{K}_{MR}}} - \mathcal{K}^{\mathbb{M}}_{MR}|_{\mathcal{O}_{\mathcal{K}_{MR}}}$ is linear, we say that there is a pure scaling discrepancy between observable model-implied and arbitrage-free CGFs.*

In summary, a discrepancy between observable arbitrage-free and model-implied CGFs can emerge either from a scaling discrepancy or from a dispersion or excess dispersion violation. Dispersion violations in the marginal CGF of returns can in principle be corrected quite precisely, when the return distribution is observable with sufficient accuracy. In contrast, correcting dispersion violations in the marginal CGF of the pricing kernel is more challenging, because dispersion is only sparsely observable along that dimension.

2.4.3 Implications of Constraints of Type (1) and Upper Arbitrage Free CGF

When informative arbitrage-free dispersion is unobservable, dispersion constraints of Type (1) imply testable constraints for the arbitrage-free CGF on the convex hull of all observable points, defined by:

$$\overline{\mathcal{O}}_{\mathcal{K}_{MR}} := \{E_\pi[(m,r)] : \text{ prior } \pi \text{ has support in } \mathcal{O}_{\mathcal{K}_{MR}}\} . \tag{2.4.6}$$

In this way, the arbitrage-free CGF is restricted also in not directly observable regions of its domain. Positivity of Jensen's gap yields for any $(m_\star, r_\star) \in \overline{\mathcal{O}}_{\mathcal{K}_{MR}}$ the upper bound:

$$\mathcal{K}_{MR}(m_\star, r_\star) \leq E_\pi[\mathcal{K}_{MR}(m,r)] , \tag{2.4.7}$$

where the right hand side of this inequality is observable because π has support on $\mathcal{O}_{\mathcal{K}_{MR}}$. The tightest such upper bound follows from the infimum over all priors with support in $\mathcal{O}_{\mathcal{K}_{MR}}$ and such that $(m_\star, r_\star) = E_\pi[(m,r)]$. This motivates the concept of an upper arbitrage-free CGF.

Definition 11 (Upper Arbitrage Free CGF). *The upper arbitrage free CGF is the function* $\mathcal{K}^U_{MR} : \mathbb{R}^{d_1+d_2} \to \mathbb{R} \cup \{+\infty\}$, *defined for any* $(m_\star, r_\star) \in \mathbb{R}^{d_1+d_2}$ *by:*[14]

$$\mathcal{K}^U_{MR}(m_\star, r_\star) := \inf_{\pi} \{E_\pi[\mathcal{K}_{MR}(m, r)]\} \ , \tag{2.4.8}$$

where the infimum is over priors with support on $\mathcal{O}_{\mathcal{K}_{MR}}$ *and such that* $(m_\star, r_\star) = E_\pi[(m, r)]$.

The upper arbitrage-free CGF is a convex upper extension of $\mathcal{K}_{MR}$ to $\mathbb{R}^{d_1+d_2}$, which coincides with $\mathcal{K}_{MR}$ on the observable set $\mathcal{O}_{\mathcal{K}_{MR}}$ and defines a finite upper bound for $\mathcal{K}_{MR}$ on the convex hull $\overline{\mathcal{O}}_{\mathcal{K}_{MR}}$.[15] By definition, $\mathcal{K}^U_{MR}$ is computable from the set $\mathcal{O}_{\mathcal{K}_{MR}}$ of observable points and implies for any $(m_\star, r_\star) \in \overline{\mathcal{O}}_{\mathcal{K}_{MR}}$ a nontrivial inequality for the specification of an arbitrage-free CGFs:

$$\mathcal{K}_{MR}(m_\star, r_\star) \leq \mathcal{K}^U_{MR}(m_\star, r_\star) \ . \tag{2.4.9}$$

Figure 2.4a illustrates the convex domain $\overline{\mathcal{O}}_{\mathcal{K}_{MR}}$ of finite upper arbitrage-free CGF values, generated by the following empirically observable CGF points: $\mathcal{K}_{MR}(1, 0) = \log B$, $\mathcal{K}_{MR}(1, 1) = 0$ and $\mathcal{K}_{MR}(0, r) = \log E[R^r]$ for $r \in (0, 1)$. $\overline{\mathcal{O}}_{\mathcal{K}_{MR}}$ is convex and closed. Outside this region, the upper bound on $\mathcal{K}_{MR}$ generated by Type (1) dispersion constraints is trivial. Naturally, a violation of bound (2.4.9) in regions where it is non-trivial provides useful information for the specification of asset pricing models.

Definition 12. *Given a model* $\mathbb{M}$ *and unobservable point* $(m_\star, r_\star) \in \overline{\mathcal{O}}_{\mathcal{K}_{MR}} \backslash \mathcal{O}_{\mathcal{K}_{MR}}$, *an arbitrage-free dispersion violation of Type (1) arises whenever* $\mathcal{K}^{\mathbb{M}}_{MR}(m_\star, r_\star) > \mathcal{K}^U_{MR}(m_\star, r_\star)$.

2.4.4 Implications of Constraints of Type (2) and Lower Arbitrage Free CGF

Given an informative unobservable arbitrage-free dispersion, dispersion constraints of Type (2) imply a second set of observable constraints for an arbitrage-free CGF:

$$\mathcal{K}_{MR}(E_\pi[(m, r)]) \leq E_\pi[\mathcal{K}_{MR}(m, r)] \ , \tag{2.4.10}$$

where the left hand side of this inequality is observable when $E_\pi[(m, r)] \in \mathcal{O}_{\mathcal{K}_{MR}}$. Inequality (2.4.10) is a lower bound for the expected arbitrage-free CGF under any prior π with observable expectation $E_\pi[(m, r)]$. From inequality (2.4.10), we obtain directly computable lower bounds for the arbitrage-free CGF in unobservable regions of its domain. Given unobservable point $(m_\star, r_\star)$, consider a prior π with support in $\mathcal{O}_{\mathcal{K}_{MR}} \cup \{(m_\star, r_\star)\}$ and such that $E_\pi[(m, r)] \in \mathcal{O}_{\mathcal{K}_{MR}}$. Inequality (2.4.10) then gives:

$$\mathcal{K}_{MR}(E_\pi[(m, r)]) \leq E_\pi[\mathcal{K}_{MR}(m, r) 1_{\mathcal{O}_{\mathcal{K}_{MR}}}(m, r)] + \pi(m_\star, r_\star)\mathcal{K}_{MR}(m_\star, r_\star) \ , \tag{2.4.11}$$

[14] By definition, $\inf_\emptyset E_\pi[\mathcal{K}_{MR}(m, r)] := +\infty$.

[15] See Peters and Wakker (1987), among other, for the properties of finite convex extensions of a convex function.

where $1_{\mathcal{O}_{\mathcal{K}_{MR}}}$ is the indicator function of set $\mathcal{O}_{\mathcal{K}_{MR}}$. In (2.4.11) all quantities but $\mathcal{K}_{MR}(m_\star, r_\star)$ are observable and the following lower bound holds:

$$\mathcal{K}_{MR}(m_\star, r_\star) \geq \frac{\mathcal{K}_{MR}(E_\pi[(m,r)]) - E_\pi[\mathcal{K}_{MR}(m,r)1_{\mathcal{O}_{\mathcal{K}_{MR}}}(m,r)]}{\pi(m_\star, r_\star)} . \tag{2.4.12}$$

The tightest such lower bound is given by the supremum of the right hand side of inequality (2.4.12) over priors with support in $\mathcal{O}_{\mathcal{K}_{MR}} \cup \{(m_\star, r_\star)\}$ and such that $E_\pi[(m,r)] \in \mathcal{O}_{\mathcal{K}_{MR}}$. This motivates the concept of a lower arbitrage-free CGF.

Definition 13 (Lower Arbitrage Free CGF)**.** *The lower arbitrage-free CGF is the function* $\mathcal{K}^L_{MR} : \mathbb{R}^{d_1+d_2} \to \mathbb{R} \cup \{-\infty\}$*, defined for any* $(m_\star, r_\star) \in \mathbb{R}^{d_1+d_2}$ *by:*[16]

$$\mathcal{K}^L_{MR}(m_\star, r_\star) := \sup_\pi \left\{ \frac{\mathcal{K}_{MR}(E_\pi[(m,r)]) - E_\pi[\mathcal{K}_{MR}(m,r)1_{\mathcal{O}_{\mathcal{K}_{MR}}}(m,r)]}{\pi(m_\star, r_\star)} \right\} , \tag{2.4.13}$$

where the supremum is over priors with support in $\mathcal{O}_{\mathcal{K}_{MR}} \cup \{(m_\star, r_\star)\}$ *and such that* $E_\pi[(m,r)] \in \mathcal{O}_{\mathcal{K}_{MR}}$.

The lower arbitrage-free CGF is a convex lower extension of $\mathcal{K}_{MR}$ to $\mathbb{R}^{d_1+d_2}$, which coincides with $\mathcal{K}_{MR}$ on the observable set $\mathcal{O}_{\mathcal{K}_{MR}}$. By construction, $\mathcal{K}^L_{MR}$ is observable from the set $\mathcal{O}_{\mathcal{K}_{MR}}$ of empirically observable points and implies a non-trivial inequality on the arbitrage-free CGF:

$$\mathcal{K}_{MR}(m_\star, r_\star) \geq \mathcal{K}^L_{MR}(m_\star, r_\star) . \tag{2.4.14}$$

This lower bound is finite whenever a prior exists with support in $\mathcal{O}_{\mathcal{K}_{MR}} \cup \{(m_\star, r_\star)\}$ and such that $E_\pi[(m,r)] \in \mathcal{O}_{\mathcal{K}_{MR}}$[17]. The set of points where $\mathcal{K}^L_{MR}$ is finite follows from the following identity:

$$(m_\star, r_\star) = (E_\pi[(m,r)] - E_\pi[(m,r)1_{\mathcal{O}_{MR}}(m,r)])/\pi(m_\star, r_\star) . \tag{2.4.15}$$

We denote by $\underline{\mathcal{O}}_{\mathcal{K}_{MR}} \subset \mathbb{R}^{d_1+d_2}$ the domain on which $\mathcal{K}^L_{MR}$ is finite. Figure 2.4b illustrates the non-convex domain $\underline{\mathcal{O}}_{\mathcal{K}_{MR}}$ of finite lower arbitrage-free CGF values, generated by the following observable CGF points: $\mathcal{K}_{MR}(1,0) = \log B$, $\mathcal{K}_{MR}(1,1) = 0$ and $\mathcal{K}_{MR}(0,r) = \log E[R^r]$ for $r \in (0,1)$. $\underline{\mathcal{O}}_{\mathcal{K}_{MR}}$ is not convex, but is closed. Outside this region, the lower bound on $\mathcal{K}_{MR}$ generated by Type (2) dispersion constraints is trivial. A violation of a nontrivial bound (2.4.14) thus provides additional information for the specification of asset pricing models.

Definition 14. *Given a model* $\mathbb{M}$ *and unobservable point* $(m_\star, r_\star) \in \underline{\mathcal{O}}_{\mathcal{K}_{MR}} \backslash \mathcal{O}_{\mathcal{K}_{MR}}$*, an arbitrage-free dispersion violation of Type (2) arises whenever* $\mathcal{K}^{\mathbb{M}}_{MR}(m_\star, r_\star) < \mathcal{K}^L_{MR}(m_\star, r_\star)$.

[16] By definition, $\sup_\emptyset E_\pi[\mathcal{K}_{MR}(m,r)] := -\infty$.

[17] $\mathcal{K}^L_{MR}$ is also related to the minimal convex extension of convex function $\mathcal{K}_{MR}$; see Dragomirescu and Ivan (1992), among others.

2.4.5 Implications of Dispersion Constrains for Dispersion Bounds

Violations of Type (1) and (2) can be generated both by an inappropriate model dispersion or an inappropriate model scaling of pricing kernel or returns. Often, scale-invariant dispersion bounds easily follow from the upper and lower CGF's in Definition 6. Indeed, for any two priors π, π', it directly follows from the definitions:

$$\mathcal{D}_\pi(M,R) \geq \left(\frac{E_\pi[\mathcal{K}^L_{MR}(m,r)] - \mathcal{K}^U_{MR}(E_\pi[(m,r)])}{tr(Var_\pi(m,r))}\right)^+ =: \mathcal{D}^L_\pi(M,R) \quad (2.4.16)$$

$$\mathcal{D}_{\pi'}(M,R) \leq \frac{E_{\pi'}[\mathcal{K}^U_{MR}(m,r)] - \mathcal{K}^L_{MR}(E_{\pi'}[(m,r)])}{tr(Var_{\pi'}(m,r))} =: \mathcal{D}^U_{\pi'}(M,R)\,, \quad (2.4.17)$$

where $(x)^+ := \max(x,0)$ is the positive part of x. Dispersion bounds (2.4.16) and (2.4.17) are not binding when the right hand side equals 0 and $+\infty$, respectively. A necessary condition for a binding bound (2.4.16) is that π has support in $\underline{\mathcal{O}}_{\mathcal{K}_{MR}}$ and $E_\pi[(m,r)] \in \overline{\mathcal{O}}_{\mathcal{K}_{MR}}$. If π has support in $\mathcal{O}_{\mathcal{K}_{MR}}$, this bound directly reflects Type (1) constraints. Whenever two such priors with identical mean exist, then dispersion bound (2.4.16) is always binding. Bound (2.4.17) is binding if and only if prior π' has support in $\overline{\mathcal{O}}_{\mathcal{K}_{MR}}$ and $E_{\pi'}[(m,r)] \in \underline{\mathcal{O}}_{\mathcal{K}_{MR}}$.[18]

2.4.6 Diagnostic Tests of Asset Pricing Models

A general approach for testing asset pricing models can rely on a test of the null hypothesis:

$$\mathcal{H}_0(m^\star, r^\star) \,:\, \mathcal{K}^L_{MR}(m^\star,r^\star) \leq \mathcal{K}^{\mathbb{M}}_{MR}(m^\star,r^\star) \leq \mathcal{K}^U_{MR}(m^\star,r^\star)\,, \quad (2.4.18)$$

over a range of relevant arguments $(m^\star, r^\star) \in \mathbb{R}^{d_1+d_2}$. On observable set $\mathcal{O}_{\mathcal{K}_{MR}}$, these inequalities are equalities and the relevant (composite) null hypothesis is:

$$\mathcal{H}_0(\mathcal{O}_{\mathcal{K}_{MR}}) \,:\, \mathcal{K}_{MR}|_{\mathcal{O}_{\mathcal{K}_{MR}}} = \mathcal{K}^{\mathbb{M}}_{MR}|_{\mathcal{O}_{\mathcal{K}_{MR}}}\,. \quad (2.4.19)$$

On unobservable set $\mathcal{O}^c_{\mathcal{K}_{MR}}$ null hypothesis (2.4.18) depends on true inequalities. Consistently with the discussion in Section 2.4.3, the inequality on the right hand side of this null hypothesis is binding on set $\overline{\mathcal{O}}_{\mathcal{K}_{MR}}$ and the resulting (composite) null hypothesis is:

$$\mathcal{H}_0(\overline{\mathcal{O}}_{\mathcal{K}_{MR}}) \,:\, \mathcal{K}^{\mathbb{M}}_{MR}|_{\overline{\mathcal{O}}_{\mathcal{K}_{MR}}} \leq \mathcal{K}_{MR}|_{\overline{\mathcal{O}}_{\mathcal{K}_{MR}}}\,. \quad (2.4.20)$$

[18] In general, bounds (2.4.16) and (2.4.17) are not both always binding for the same prior π: $\mathcal{D}^L_\pi(M,R) \leq \mathcal{D}_\pi(M,R) \leq \mathcal{D}^U_\pi(M,R)$. A necessary condition is that π has support in $\underline{\mathcal{O}}_{\mathcal{K}_{MR}} \cap \overline{\mathcal{O}}_{\mathcal{K}_{MR}}$ and is such that $E_\pi[(m,r)] \in \underline{\mathcal{O}}_{\mathcal{K}_{MR}} \cap \overline{\mathcal{O}}_{\mathcal{K}_{MR}}$. Such a situation can arise when the observable set $\mathcal{O}_{\mathcal{K}_{MR}}$ is not extremal.

Similarly, the inequality on the left hand side of null hypothesis (2.4.18) is binding on set $\underline{\mathcal{O}}_{\mathcal{K}_{MR}}$ and the resulting (composite) null hypothesis is:

$$\mathcal{H}_0(\underline{\mathcal{O}}_{\mathcal{K}_{MR}}) \; : \; \mathcal{K}_{MR}|_{\underline{\mathcal{O}}_{\mathcal{K}_{MR}}} \leq \mathcal{K}^{\mathbb{M}}_{MR}|_{\underline{\mathcal{O}}_{\mathcal{K}_{MR}}} \; . \tag{2.4.21}$$

Model Diagnostics Based on Null Hypothesis $\mathcal{H}_0(\mathcal{O}_{\mathcal{K}_{MR}})$

A diagnostics test of null hypothesis

$$\mathcal{H}_0(\mathcal{O}_{\mathcal{K}_{MR}}) = \bigcap_{(m^\star,r^\star)\in\mathcal{O}_{\mathcal{K}_{MR}}} \mathcal{H}_0(m^\star, r^\star) \; , \tag{2.4.22}$$

tests the specification of a CGF in observable parts of the domain of the arbitrage-free CGF. Such a test is naturally complemented by a test of the observable model-implied dispersion and excess dispersion properties, which is a test of the (composite) null hypothesis:

$$\mathcal{J}^{\mathbb{M}}_{\pi}(M,R) \;\; = \;\; \mathcal{J}_{\pi}(M,R) \; , \tag{2.4.23}$$

$$\Delta\mathcal{J}^{\mathbb{M}}_{\pi_1,\pi_2}(M,R) \;\; = \;\; \Delta\mathcal{J}_{\pi_1,\pi_2}(M,R) \; , \tag{2.4.24}$$

for any observable arbitrage-free dispersion and excess dispersion. Consistent with Proposition 2, this two-step approach isolates potential rejections of $\mathcal{H}_0(\mathcal{O}_{\mathcal{K}_{MR}})$ due to an observable scaling mismatch from those due to an inappropriate observable model dispersion or excess dispersion.

Model Diagnostics Based on Null Hypotheses $\mathcal{H}_0(\overline{\mathcal{O}}_{\mathcal{K}_{MR}})$ and $\mathcal{H}_0(\underline{\mathcal{O}}_{\mathcal{K}_{MR}})$

A diagnostics test of null hypotheses

$$\mathcal{H}_0(\overline{\mathcal{O}}_{\mathcal{K}_{MR}}) = \bigcap_{(m^\star,r^\star)\in\overline{\mathcal{O}}_{\mathcal{K}_{MR}}} \mathcal{H}_0(m^\star, r^\star) \;\; ; \;\; \mathcal{H}_0(\underline{\mathcal{O}}_{\mathcal{K}_{MR}}) = \bigcap_{(m^\star,r^\star)\in\underline{\mathcal{O}}_{\mathcal{K}_{MR}}} \mathcal{H}_0(m^\star, r^\star) \tag{2.4.25}$$

tests the specification of an asset pricing model in unobservable parts of the domain of the arbitrage-free CGF. In order to isolate a potential violation of $\mathcal{H}_0(\overline{\mathcal{O}}_{\mathcal{K}_{MR}})$ or $\mathcal{H}_0(\underline{\mathcal{O}}_{\mathcal{K}_{MR}})$ due to inappropriate scaling from those due to inappropriate unobservable dispersion or excess disperion, it is convenient to complement these tests by a set of scale invariant dispersion tests, based on the dispersion bounds introduced in Section 2.4.5. Using dispersion bounds, a scale independent diagnostics test for asset pricing models can rely on a test of the inequalities:

$$\mathcal{D}^{L}_{\pi}(M,R) \leq \mathcal{D}^{\mathbb{M}}_{\pi}(M,R) \leq \mathcal{D}^{U}_{\pi}(M,R) \; , \tag{2.4.26}$$

over a range of relevant priors π implying a binding dispersion bound. When two priors π_1, π_2 with support in $\mathcal{O}_{\mathcal{K}_{MR}}$ exist, such that $E_{\pi_1}[(m,r)] = E_{\pi_2}[(m,r)]$ and $E_{\pi_1}[\mathcal{K}_{MR}(m,r)] > E_{\pi_2}[\mathcal{K}_{MR}(m,r)]$, a binding lower bound in the LHS of inequality

(2.4.26) is easily available, using the convexity of $\mathcal{K}^U_{MR}$:

$$\mathcal{D}^L_{\pi_1}(M,R) \geq \frac{E_{\pi_1}[\mathcal{K}_{MR}(m,r)] - \mathcal{K}^U_{MR}(E_{\pi_1}[(m,r)])}{tr(Var_{\pi_1}(m))} \geq \frac{E_{\pi_1}[\mathcal{K}_{MR}(m,r)] - E_{\pi_2}[\mathcal{K}_{MR}(m,r)]}{tr(Var_{\pi_1}(m))} > 0\,.$$

In contrast, a binding upper bound in the RHS of inequalities (2.4.26) emerges only when observable set $\mathcal{O}_{\mathcal{K}_{MR}}$ is not extremal. The observable set implied by the Bansal and Yaron, 2004 long-run risk model introduced in Section 2.6 is extremal. Therefore, our empirical analysis of the arbitrage free dispersion properties of long-run risk models in Section 2.6 is based on tests of lower dispersion bounds.

2.5 Explicit Pricing Kernel Bounds Induced by Dispersion Constraints

We derive with a unifying dispersion approach explicit model-free pricing kernel bounds induced by constrainst of Type (1) or (2). This approach allows an easy derivation of existing sharp univariate bounds in the literature. Moreover, it is suitable for a natural extension to economies with multiple pricing kernel components, for which we obtain new sharp pricing kernel bounds. We illustrate our approach in the benchmark economy with a single pricing kernel. We then extend it to pricing kernels with transienst and persistent components and to international economies with domestic and foreign pricing kernels.

2.5.1 Univariate Pricing Kernel Bounds

Given a univariate pricing kernel ($d_1 = 1$) and d_2 risky returns, we consider an arbitrage-free joint CGF restricted by complete return observability ($(0,r) \in \mathcal{O}_{\mathcal{K}_{MR}}$ for any $r \in \mathbb{R}^{d_2}$), a risk-free bond with price B ($(1,0_{d_2}) \in \mathcal{O}_{\mathcal{K}_{MR}}$) and risky returns $R = (R_1,\ldots,R_{d_2})$ ($(1,e_i) \in \mathcal{O}_{\mathcal{K}_{MR}}$ for each unit vector e_i in $\mathbb{R}^{d_2}$).

Dispersion Constraints of Type (1) and Entropy Bounds

For any $\alpha \in (0,1)$, a first set of dispersion constraints of Type (1) follows using a Bernoulli prior π with mass $\alpha \in (0,1)$ on $(1,e_i)$ and mass $1-\alpha$ on $(0,-\frac{\alpha}{1-\alpha}e_i)$. Indeed, from the dispersion constraint

$$\mathcal{J}_\pi(M,R) = E_\pi[\mathcal{K}_{MR}(m,r)] - \mathcal{K}_{MR}(E_\pi[m,r]) \geq 0\,, \tag{2.5.1}$$

we have $\mathcal{K}_{MR}(1,e_i) = 0$ and $(0,-\frac{\alpha}{1-\alpha}e_i) \in \mathcal{O}_{\mathcal{K}_{MR}}$, i.e., prior π has support in $\mathcal{O}_{\mathcal{K}_{MR}}$. Consequently, inequality (2.5.1) defines an arbitrage-free dispersion constraint of Type (1). Explicit calculations give:

$$\frac{1}{1-\alpha}\log E[M^\alpha] = \frac{1}{1-\alpha}\mathcal{K}_{MR}(\alpha,0_{d_2}) \leq \mathcal{K}_{MR}(0,-\frac{\alpha}{1-\alpha}e_i) = \log E[R_i^{-\alpha/(1-\alpha)}] \tag{2.5.2}$$

Equivalently, this is a lower bound on the Rényi, 1960 entropy of the pricing kernel:

$$\mathcal{E}_\alpha(M) = \frac{1}{\alpha(\alpha-1)} \log E[(M/E(M))^\alpha] \geq -\frac{1}{\alpha} \log E[R_i^{-\alpha/(1-\alpha)}] - \frac{1}{\alpha-1} \log E[M] \quad (2.5.3)$$

As $\alpha \to 0$, this is the entropy bound in, e.g., Alvarez and Jermann, 2005: $\mathcal{E}_0(M) \geq \log E[R_i E(M)]$.

Dispersion Constraints of Type (2) and Entropy Bounds

For any $\alpha > 1$, a dispersion constraint of Type (2) follows using a prior with mass $1/\alpha \in (0,1)$ in $(\alpha, 0_{d_2})$ and mass $(\alpha-1)/\alpha$ in $(0, \frac{\alpha}{\alpha-1} e_i)$. Indeed, since $E_\pi[m,r] = (1, e_i) \in \mathcal{O}_{\mathcal{K}_{MR}}$, we obtain:

$$0 = \mathcal{K}_{MR}(1, e_i) \leq \frac{1}{\alpha} \mathcal{K}_{MR}(\alpha, 0_{d_2}) + \frac{\alpha-1}{\alpha} \mathcal{K}_{MR}\left(0, -\frac{\alpha}{1-\alpha} e_i\right) , \quad (2.5.4)$$

which is pricing kernel bound (2.5.2) for $\alpha > 1$. Such pricing kernel bounds for $\alpha > 0$ are equivalent to the pricing kernel bounds derived in Liu (2015) using Hölder-type inequalities. For $\alpha < 0$, we obtain a second set of constraints of Type (2) using a prior with mass $1/(1-\alpha) \in (0,1)$ in $(\alpha, 0_{d_2})$ and mass $-\alpha/(1-\alpha)$ in $(1, e_i)$. Indeed, as $E_\pi[m,r] = (0, -\frac{\alpha}{1-\alpha} e_i) \in \mathcal{O}_{\mathcal{K}_{MR}}$ and $\mathcal{K}_{MR}(1, e_i) = 0$, we have:

$$\mathcal{K}_{MR}(0, -\frac{\alpha}{1-\alpha} e_i) \leq \frac{1}{1-\alpha} \mathcal{K}_{MR}(\alpha, 0_{d_2}) , \quad (2.5.5)$$

which is the reversed pricing kernel bound (2.5.2) for $\alpha < 0$. Equivalently,

$$\mathcal{E}_\alpha(M) \leq \frac{(1-\alpha) \log E[R_i^{-\alpha/(1-\alpha)}] - \alpha \log E[M]}{\alpha(\alpha-1)} . \quad (2.5.6)$$

These bounds for $\alpha < 0$ are equivalent to the bounds derived in Snow (1991) using Hölder-type inequalities. In summary, we have obtained the following proposition.

Proposition 3. *For any $\alpha \in \mathbb{R}$, the following dispersion constraints hold:*

$$\frac{\mathcal{K}_M(\alpha)}{\alpha(\alpha-1)} \geq -\frac{\mathcal{K}_{R_i}(\alpha/(\alpha-1))}{\alpha} \; ; \; i = 1, \ldots, d_2 . \quad (2.5.7)$$

Figure 2.5 illustrates for $d_1 = d_2 = 1$ the construction of the above pricing kernel bounds. For $\alpha = 1/2$, we appy a constraint of Type (1) with observable points $(m,r) = (1,1)$ and $(m,r) = (0,-1)$. In this case, the unobserved point $(m,r) = (0, 1/2)$ lies in the convex hull of the observable points. For $\alpha = 2$ ($\alpha = -1$), we apply a constraint of Type (2) with observable points $(m,r) = (1,1)$ and $(m,r) = (0,2)$ (points $(m,r) = (1,1)$ and $(m,r) = (0,1/2)$). In these last two cases, the unobserved point $(m,r) = (0,2)$ ($(m,r) = (0,-1)$) lies outside of the convex hull of the observable points.

Bound Tightness

An important question is whether the pricing kernel bounds resulting from the dispersion constaints in Proposition 3 are tight, in the sense that they are the sharpest bounds implied by the arbitrage-free constraints on returns $(R_0, R_1, \ldots, R_{d_2})$, where R_0 is the risk-free return.[19] Given the joint distribution of returns, the tightest lower bound on $\frac{\mathcal{K}_M(\alpha)}{\alpha(\alpha-1)}$ is the one implied by the minimum divergence pricing kernel in Almeida and Garcia (2017).[20] We find that the pricing kernel bounds implied by arbitrage-free dispersion constraints are sharp, after a maximization of the right hand side on inequality (2.5.7) over all returns of portfolios with weights $1-\sum_{i=1}^{d_2}\lambda_i, \lambda_1, \ldots, \lambda_{d_2}$ in returns $R_0, \ldots, R_{d_2}$.

Proposition 4 (Bound Tightness). *Let $M^\star$ be the solution of the minimal divergence problem in Almeida and Garcia (2017):*

$$\inf_M \left\{ \frac{\mathcal{K}_M(\alpha)}{\alpha(\alpha-1)} \right\} , \tag{2.5.10}$$

subject to $M > 0$ and $\mathcal{K}_{MR}(1, e_i) = 0$ for $i = 0, \ldots, d_2$. Consider the following maximization problem:

$$\sup_\lambda \left\{ -\frac{\mathcal{K}_{R_\lambda}(\alpha/(\alpha-1))}{\alpha} \right\} , \tag{2.5.11}$$

where $R_\lambda = \sum_{i=1}^{d_2}\lambda_i R_i + (1-\sum_{i=1}^{d_2}\lambda_i)R_0$. (with the constraint $R_\lambda > 0$). Given the solution $\lambda^\star$ to this problem, it follows:

$$\frac{\mathcal{K}_{M^\star}(\alpha)}{\alpha(\alpha-1)} = -\frac{\mathcal{K}_{R_{\lambda^\star}}(\alpha/(\alpha-1))}{\alpha} , \tag{2.5.12}$$

and the minimal divergence stochastic discount factor is given by $M^\star = R_{\lambda^\star}^{-1/(1-\alpha)}/E[R_{\lambda^\star}^{-\alpha/(1-\alpha)}]$.

Proposition 4 shows that the tightest lower bound on $\frac{\mathcal{K}_M(\alpha)}{\alpha(\alpha-1)}$, which is compatible with a stochastic discount factor pricing returns $R_0, \ldots, R_{d_2}$, follows from a single

[19] Bansal and Lehmann (1997) and Alvarez and Jermann (2005), among others, show that the tightest pricing kernel entropy bound, which is obtained for $\alpha \to 0$ in our setting, is the one generated by the return of the growth optimal portfolio. By construction, that bound is equivalent to the bound generated for $\alpha = 0$ by arbitrage-free dispersion constraints incorporating the observability of the return on the growth optimal portfolio.

[20] This follows from the equivalence of the optimization problem:

$$\inf_M \left\{ \frac{\mathcal{K}_M(\alpha)}{\alpha(\alpha-1)} \right\} \; s.t. \; \mathcal{K}_{MR}(1, e_i) = 0 \; (i = 0, \ldots, d_2) , \tag{2.5.8}$$

with the minimal divergence problem:

$$\inf_M \left\{ \frac{E[M^\alpha] - E[M]^\alpha}{\alpha(\alpha-1)} \right\} \; s.t. \; E[MR_i] = 1 \; (i = 0, \ldots, d_2) , \tag{2.5.9}$$

in Almeida and Garcia (2017).

arbitrage-free dispersion constraint for portfolio return $R_{\lambda^\star}$.[21] The bound tightness in Proposition 4 also implies closed-form upper and lower arbitrage free CGF of the pricing kernel, wich explicit domains $D_U = [0,1]$ and $D_L = [0,1]^c$ where these functions take finite values:

$$\mathcal{K}_M(\alpha) \geq \mathcal{K}_M^L(\alpha) = (1-\alpha)\mathcal{K}_R(\alpha/(\alpha-1)) > -\infty \; ; \; \alpha \in D_L \; , \tag{2.5.13}$$

$$\mathcal{K}_M(\alpha) \leq \mathcal{K}_M^U(\alpha) = (1-\alpha)\mathcal{K}_R(\alpha/(\alpha-1)) < \infty \; ; \; \alpha \in D_U \; . \tag{2.5.14}$$

2.5.2 Multivariate Pricing Kernel Bounds

Using the same general dispersion approach as in the previous section, we now address multivariate settings with multiple pricing kernel components.

Dispersion Constraints on Transient and Persistent Pricing Kernel Components

In the context of Section 2.2.4, we directly obtain a family of dispersion constraints of Type (1), using for given $0 < \beta < 1$ a prior with mass β in $(m,r) = (1,1,1) \in \mathcal{O}_{\mathcal{K}_{MR}}$ and mass $1-\beta$ in $(m,r) = (-(\beta-\alpha), 0, -\beta)/(1-\beta) \in \mathcal{O}_{\mathcal{K}_{MR}}$. In this way, we have for any $\alpha \in \mathbb{R}$:

$$\frac{\mathcal{K}_{M^T M^P}(\alpha,\beta)}{\beta(\beta-1)} \geq -\frac{\mathcal{K}_{M^T R}\left(-\frac{\beta-\alpha}{1-\beta}, -\frac{\beta}{1-\beta}\right)}{\beta} = -\frac{\mathcal{K}_{R_\infty R}\left(\frac{\beta-\alpha}{1-\beta}, -\frac{\beta}{1-\beta}\right)}{\beta} \; . \tag{2.5.15}$$

This bound constraints the joint distribution of transient and persistent pricing kernel components and coincides with the univariate pricing kernel bound (2.5.7) when $\beta = \alpha$. Note that a violation of bound (2.5.15) can arise from an inappropriate model-implied scaling of $M^T = 1/R_\infty$. A scale-independent bound equivalent to bound (2.5.15) is given in the next proposition, proven in the Supplemental Appendix.

Proposition 5 (Bound Tightness with SDF Decomposition)**.** *Given the physical distribution $\mathbb{P}$ of pricing kernel components and returns, let equivalent measure $\mathbb{T}$ be defined by the Radon-Nikodym derivative $\frac{d\mathbb{T}}{d\mathbb{P}} := \frac{(M^T)^\gamma}{E[(M^T)^\gamma]}$ for some $\gamma \in \mathbb{R}$.[22] Then $M^{\mathbb{T}} := M(M^T)^{-\gamma}E[(M^T)^\gamma]$ is a stochastic discount vector with respect to measure $\mathbb{T}$ and the following bound is sharp for any $\beta \in \mathbb{R}$:*

$$\frac{\mathcal{K}_{M^{\mathbb{T}}}^{\mathbb{T}}(\beta)}{\beta(\beta-1)} \geq -\frac{\mathcal{K}_R^{\mathbb{T}}(-\beta/(1-\beta))}{\beta} \; . \tag{2.5.16}$$

[21] Proposition 4 also implies that while the pricing kernel bounds in Snow (1991) and Liu (2015) derived from univariate pricing constraints are not sharp in general, they are after an optimization with respect to the family of portfolio returns generated by the priced underlying assets in an arbitrage free market. This follows directly from the equivalence of the minimum divergence stochastic discount factor bound in Almeida and Garcia (2017) and the optimized dispersion bound in Proposition 4.

[22] This change of measure is well-defined if and only if the marginal CGF of M^T in γ is well-defined.

This bound is equivalent to the bound

$$\frac{\mathcal{K}_{M^T M^P}(\gamma + (1-\gamma)\beta, \beta)}{\beta(\beta - 1)} \geq -\frac{\mathcal{K}_{M^T R}(\gamma, -\beta/(1-\beta))}{\beta} , \tag{2.5.17}$$

in the sense that the difference of the LHS and the RHS of inequalities (2.5.16) and (2.5.17) is identical.

The parameter choice $\gamma = \frac{\alpha-\beta}{1-\beta}$ in Proposition 5 implies bound (2.5.15) for any $\alpha, \beta \in \mathbb{R}$ such that $\mathcal{K}_M((\alpha-\beta)/(1-\beta))$ is well-defined. Note while bound (2.5.16) is equivalent to bound (2.5.17), it is also robust to the scale of M^T. In this sense, a violation of bound (2.5.16) by an asset pricing model has the desirable property of being robust to an inappropriate model-implied scaling of M^T, deriving, e.g., from an inappropriate empirical measurement of long term real bond returns.

Chernoff, 1952 Entropy Bounds on Domestic and Foreign Pricing Kernels

In the context of Section 2.3.2, inequality $\min(x, y) \leq x^\alpha y^{1-\alpha}$ yields for any $\alpha \in (0, 1)$ the following Chernoff, 1952 information bound on the average minimal pricing kernel:[23]

$$E[\min(M_d/E[M_d], M_f/E[M_f])] \leq \exp(-\mathcal{CI}_*(M_d, M_f)) . \tag{2.5.18}$$

When markets are complete, the forward exchange rate return $F_e = (M_f/E[M_f])/(M_d/E[M_d])$ implies:

$$E\left[\frac{M_d}{E[M_d]} \max(0, 1 - F_e)\right] \geq 1 - \exp(-\mathcal{CI}_*(M_d, M_f)) \approx \mathcal{CI}_*(M_d, M_f) , \tag{2.5.19}$$

i.e., the (forward) price of an at-the-money put option on the (forward) exchange rate is a tight upper bound on the Chernoff information of domestic and foreign pricing kernels.[24]

Dispersion Constraints of Type (1) on Domestic and Foreign Pricing Kernels

In a $d-$country economy with pricing kernel components $M = (M_1, \ldots, M_d)$ pricing the gross returns $R = (R_1, \ldots, R_d)$, the following pricing constraints hold for $i = 1, \ldots, d$:

$$\mathcal{K}_{M_i R_i}(1, 1) = \mathcal{K}_{MR}([e_i', e_i']) = 0 . \tag{2.5.20}$$

[23] Recall Definition (2.3.13) of Chernoff information. We make use of inequality $\min(x, y) \leq x^\alpha y^{1-\alpha}$ for $x, y \geq 0$ and $\alpha \in (0, 1)$.

[24] From the symmetry of Chernoff information, the same bound applies for the (forward) price of an at-the-money put option on the (forward) exchange return $1/F_e$.

Given strictly positive vector $\alpha = (\alpha_1, \ldots, \alpha_d)$ such that $||\alpha||_1 := \sum_{i=1}^d \alpha_i < 1$, we also have:

$$[\alpha, 0_d] = \left(1 - \sum_{i=1}^d \alpha_i\right) \left[0_d, -\frac{\alpha}{1 - \sum_{i=1}^d \alpha_i}\right] + \alpha_i \sum_{i=1}^d [e_i, e_i] \,. \tag{2.5.21}$$

Dispersion constraints of Type (1) then directly imply following multivariate pricing kernel bound:

$$\frac{\mathcal{K}_M(\alpha)}{(\sum_{i=1}^d \alpha_i - 1) \prod_{i=1}^d \alpha_i} \geq -\frac{\mathcal{K}_R\left(\alpha/(\sum_{i=1}^d \alpha_i - 1)\right)}{\prod_{i=1}^d \alpha_i} \,. \tag{2.5.22}$$

This bound is a natural multivariate version of bound (2.5.2). Moreover, it can be optimally sharpened in economies with multiple domestic and foreign returns. Precisely, let pricing kernel M_i price returns $R_{i0}, \ldots R_{iN_i}$ in market $i = 1, \ldots, d$, where R_{i0} is the $i-$th risk-free return. Then, the optimization of lower bound (2.5.22) over portfolios of these returns provides in Proposition 6 the sharpest bound.[25]

Proposition 6 (Multivariate Bound Tightness)**.** *Consider for $\alpha \in \mathbb{R}^d_{++}$ such that $||\alpha||_1 < 1$ the pricing kernel vector $M^\star = (M_1^\star, \ldots, M_d^\star)$ that solves the following minimum divergence problem:*[26]

$$\inf_M \left\{ \frac{\mathcal{K}_M(\alpha)}{(\sum_{i=1}^d \alpha_i - 1) \prod_{i=1}^d \alpha_i} \right\} \,, \tag{2.5.23}$$

subject to the following moment conditions, indexed by $i = 1, \ldots, d$ and $k_i = 0, \ldots, N_i$:

$$\mathcal{K}_{M_i R_{ik_i}}(1, 1) = 0 \,. \tag{2.5.24}$$

Further, consider the solution $\lambda^\star$ of the maximization problem:

$$\sup_\lambda \left\{ -\frac{\mathcal{K}_{R_\lambda}(\alpha/(\sum_{1=1}^d \alpha_i - 1))}{\prod_{i=1}^d \alpha_i} \right\} \,, \tag{2.5.25}$$

where $R_\lambda = (R_{1\lambda_1}, \ldots, R_{d\lambda_d})$ and $R_{i\lambda_i} = \sum_{k_i=1}^{N_i} \lambda_{ik} R_{ik} + (1 - \sum_{k_i=1}^{N_i} \lambda_{ik} R_{i0})$ is the return of a portfolio of returns (with constraint $R_{i\lambda_i} > 0$) denominated in the $i-th$ domestic currency. It then follows:

$$\frac{\mathcal{K}_{M^\star}(\alpha)}{(\sum_{i=1}^d \alpha_i - 1) \prod_{i=1}^d \alpha_i} = -\frac{\mathcal{K}_{R_{\lambda^\star}}(\alpha/(1 - \sum_{1=1}^d \alpha_i))}{\prod_{i=1}^d \alpha_i} \,. \tag{2.5.26}$$

[25] The proof is collected in the Supplemental Appendix.

[26] $\mathbb{R}^d_{++}$ denotes the $d-$dimensional strictly positive cone.

The optimal pricing kernel $M^\star := (M_1^\star, \ldots, M_d^\star)$ *has components given explicitly by:*

$$M_i^\star = \frac{\left[R_{i\lambda_i^\star}^{1-\sum_{j\neq i}\alpha_j} \prod_{j\neq i}^d R_{j\lambda_j^\star}^{\alpha_j}\right]^{1/(\sum_{j=1}^d \alpha_j - 1)}}{E\left[\prod_{j=1}^d R_{j\lambda_j^\star}^{\alpha_j/(\sum_{j=1}^d \alpha_j - 1)}\right]} \; ; \; i = 1, \ldots, d \, . \tag{2.5.27}$$

From Proposition 6, the tightest lower bound on $\mathcal{K}_M(\alpha)/((\sum_{i=1}^d \alpha_i - 1)\prod_{i=1}^d \alpha_i)$, which is compatible with a $d-$dimensional vector M of pricing kernels for the given sets of returns, is obtained from a single dispersion constraint of Type (1) applied to the vector of portfolio returns $R_{\lambda^\star} = (R_{1\lambda_1^\star}, \ldots, R_{d\lambda_d^\star})$. By construction, the tightness result in Proposition 6 also identifies in closed-form the upper arbitrage-free CGF on domain $D := \{\alpha \in \mathbb{R}_{++}^d : ||\alpha||_1 < 1\}$:

$$\mathcal{K}_M(\alpha) \leq \mathcal{K}_M^U(\alpha) = \left(1 - \sum_{i=1}^d \alpha_i\right) \mathcal{K}_{R_{\lambda^\star}}\left(\alpha/(1 - \sum_{1=1}^d \alpha_i)\right) \; ; \; \alpha \in D \, . \tag{2.5.28}$$

This result also induce an obvious generalization of univariate dispersion bounds. For instance, a prior π_α with mass α_i on observable point $[e_i', 0_d']$ $(i = 1, \ldots, d)$ and mass $1 - \sum_{i=1}^d \alpha_i$ on point 0_{2d} implies the lower dispersion bound

$$\begin{aligned} \mathcal{D}_{\pi_\alpha}(M) &= \frac{E_{\pi_\alpha}[\mathcal{K}_M(m)] - \mathcal{K}_M(E_{\pi_\alpha}[m])}{\sum_{i=1}^d \alpha_i(1-\alpha_i)} \\ &\geq \frac{-\sum_{i=1}^d \alpha_i \log R_{i0} + (\sum_{i=1}^d \alpha_i - 1)\mathcal{K}_{R_{\lambda_\star}}(\alpha/(\sum_{i=1}^d \alpha_i - 1))}{\sum_{i=1}^d \alpha_i(1-\alpha_i)} \end{aligned} \tag{2.5.29}$$

which is a natural multivariate extension of the univariate entropy bound (2.5.3). This bound is computable from the marginal distribution of returns across different markets and it yields restrictions on both the marginal distribution of pricing kernel components and their joint dependence.

It is useful to recall that no assumption on market completeness has been made in Proposition 6, such as assumptions about the structure of the exchange rates between domestic and foreign markets. Obviously, the optimal bound in Proposition 6 is sharper whenever the set of returns $R_{i1}, \ldots, R_{iN_i}$ is wider in each market. Using exchange rate markets, the set of domestic asset returns is naturally extended, by adding to each set of domestic returns the set of foreign returns converted in domestic currency with the correponsing exchange rate return.

2.5.3 Relation to Other Bounds in the Literature

Bakshi and Chabi-Yo, 2012 investigate the variance of the permanent / transitory component stochastic discount factor. Clearly the variance of the permanent component of the SDF can be written as:

$$Var\left(M^P\right) = \mathbb{E}\left[(M^P)^2\right] - \mathbb{E}^2\left[M^P\right] = \exp(\kappa(2)) - 1 \tag{2.5.30}$$

which is just a monotone transformation of the divergence measure $\mathcal{D}_H$ with H binomially distributed over {0,2}. Finding the optimal value for $\mathcal{D}_H$ by optimizing over λ as in Proposition 4 provides the entropy bound given in Bakshi and Chabi-Yo, 2012. Similar conclusion holds for the transitory component as well.

Bakshi and Chabi-Yo, 2014 introduce the "entropy" of M^2 as:

$$L(M^2) = \log(\mathbb{E}(M^2)) - \mathbb{E}(\log(M^2)) = \frac{1}{4}\lim_{\alpha\to 0}\mathcal{D}_{H_\alpha} \tag{2.5.31}$$

where H_α is a binomial distribution taking values in $\{0, 2\}$ with probabilities $(1-\alpha)$ and α respectively. For each α we can optimize over λ to achieve the maximal value of $\mathcal{D}_{H_\alpha}$ given the lower bound (2.5.22). Taking the limit as $\alpha \to 0$ results in the lower bound given in Bakshi and Chabi-Yo, 2014.

2.6 Arbitrage Free Dispersion in Long Run Risk Models

We characterize the AFD of Bansal and Yaron, 2004 long-run risk (LRR) model, based on the recent model estimation in Bansal, Kiku, and Yaron, 2016, which incorporates temporal aggregation.

2.6.1 LRR Model

The LRR model is based on a representative agent with recursive preferences maximizing the life-time utility,

$$V_t = \left[(1-\delta)C_t^{\frac{1-\gamma}{\theta}} + \delta\left(E_t[V_{t+1}^{1-\gamma}]\right)^{1/\theta}\right]^{\frac{\theta}{1-\gamma}}, \tag{2.6.1}$$

where C_t is consumption at time t, $0 < \delta < 1$ the time preference rate, γ the parameter of relative risk aversion and $\theta := \frac{1-\gamma}{1-1/\psi}$, with ψ is the elasticity of intertemporal substitution (IES). Utility maximization is subjec to to the budget constraint $W_{t+1} = (W_t - C_t)R_{c,t+1}$, where $R_{c,t+1}$ is the return on invested wealth, and consumption growth Δc_{t+1} satisfies the LRR dynamics:

$$\begin{aligned} \Delta c_{t+1} &= \mu_c + x_t + \sigma_t\eta_{t+1}\,, \\ x_{t+1} &= \rho x_t + \psi_e\sigma_t e_{t+1}\,, \\ \sigma^2_{t+1} &= \sigma_0^2 + \nu(\sigma_t^2 - \sigma_0^2) + \sigma_w w_{t+1}\,. \end{aligned} \tag{2.6.2}$$

with $(\eta_{t+1}, e_{t+1}, w_{t+1}) \sim IIN(0, I_3)$. The resulting (single-period) pricing kernel is

$$M_{t,t+1} = \delta^\theta (C_{t+1}/C_t)^{-\theta/\psi} R_{c,t+1}^{\theta-1}\,. \tag{2.6.3}$$

We borrow from the headline estimated specification with monthly aggregation intervals in Bansal, Kiku, and Yaron, 2016, making use of the parameter estimates in

the right-hand-side panel of Table II in their paper, and study the joint AFD properties of transient and persistent pricing kernel components in the LRR model.[27] Bansal, Kiku, and Yaron estimate the parameters in the LRR model using annual time series of real consumption, the stock market portfolio and real risk-free rates, in the sample period from 1930 to 2009.[28] In order to factorize the pricing kernel in the LRR model into transient and persistent conponents, we follow Alvarez and Jermann, 2005 and make use of proxies based on the return of long-maturity bonds. This data is available from CRSP's Fixed Term Indices dataset at a monthly frequency. Consistently with the aggregation procedure in Bansal, Kiku, and Yaron, 2016, we aggregate these returns to an annual frequency and convert them to real returns using the CPI from the BLS.

2.6.2 Joint CGF of Transient and Persistent Pricing Kernel Components

We factorize the annual pricing kernel in the LRR model as $M_{t+1} = M^T_{t+1} M^P_{t+1}$, where permanent component M^P is such that $E_t[M^P_{t+1}] = 1$. Following Alvarez and Jermann, 2005, we identify the transient component using the annual return on the infinite maturity bond: $M^T_{t+1} = 1/R_{\infty,t+1}$. We calculate the model-implied CGF of $M = (M^T, M^P)$ by Monte Carlo simulation. Precisely, using monthly aggregation steps $s = 1, \ldots, 12$ we simulate $N = 10^6$ annual paths of state dynamics (2.6.2) in the LRR model, based on parameter estimates from Table II in Bansal, Kiku, and Yaron, 2016. Along each of the N annual simulated paths, we calculate on a monthly frequency the time series of single-period stochastic discount factors (2.6.3) and model-implied long-maturity bond prices, from which we obtain the annual pricing kernel $M_{t+1}(u) = \prod_{s=1}^{12} M_{s,s+1}(u)$ and the annual long maturity bond returns $R_{\infty,t+1}(u)$, $u = 1, \ldots, N$.[29] Given the simulated distribution of $M^T_{t+1}(u) = 1/R_{\infty,t+1}(u)$ and $M_{t+1}(u)$ realizations, we calculate the simulated realization of $M^P_{t+1}(u) = M_{t+1}(u)/M^T_{t+1}(u)$. Finally, for powers (t, p) in domain $D = (0, 1)^2$, we compute the Monte Carlo CGF estimate as:

$$\mathcal{K}^{\mathbb{M}}_M(p,t) = \log E^{\mathbb{M}}\left[\left(M^T_{t+1}\right)^t \left(M^P_{t+1}\right)^p\right] \approx \frac{1}{N}\sum_{u=1}^{N} (M^T_{t+1}(u))^t (M^P_{t+1}(u))^p \,, \qquad (2.6.4)$$

[27] This specification is not rejected by the overidentification tests in Bansal, Kiku, and Yaron, 2016.

[28] Consumption represents per-capita real consumption expenditures on nondurables and services from NIPA tables. Aggregate stock market data consist of annual observations of returns, dividends, and prices of the CRSP value-weighted portfolio of all stocks traded on the NYSE, AMEX, and NASDAQ. The ex-ante real risk-free rate is constructed from a projection of the ex-post real rate on the current nominal yield and inflation over the previous year. Market data are converted to real using the consumer price index (CPI) from the BLS.

[29] The yields of discount real bonds are affine functions of the state variables in the LRR model after a log-linearization. These affine functions can be calculated recursively as bond maturity increases. In this way, we obtain the price of the infinite maturity bond numerically for a sufficiently long maturity, avoiding to solve the eigenfunction problem implied by Perron-Frobenius theorem; see, e.g., Bakshi and Chabi-Yo, 2014. We follow Bansal, Kiku, and Yaron, 2016 and log-linearize around the mean value of the price-consumption ratio. This provides a fixed-point problem that is solved numerically.

where $\mathbb{M}$ emphasizes the model-implied character of this CGF. The CGF $\mathcal{K}_M^{\mathbb{M}}|_D$ in the LRR model is plotted in the left panel of Figure 2.6. It features a pronounced convexity along the p-axis and a much flatter profile along the t-axis. Consistent with the intuition in, e..g, Alvarez and Jermann, 2005, these convexity properties induce a significant dispersion in the permanent pricing kernel component and a much lower dispersion in the transient pricing kernel component of the LRR model.

2.6.3 Observable Set $\mathcal{O}_{\mathcal{K}_{MR}}$ for the LRR Model

Given a suitable risky portfolio return R_λ, we obtain the observable arbitrage-free CGF using following observable set $O_{\mathcal{K}_{MR_\lambda}}$.

Assumption 1 (Observable Set)**.** Set $O_{\mathcal{K}_{MR_\lambda}}$ is defined by the following observable points:

(1) Restriction at the origin: $(0,0,0) \in O_{\mathcal{K}_{MR_\lambda}}$.

(2) Martingale normalization: $(0,1,0) \in \mathcal{O}_{\mathcal{K}_{MR_\lambda}}$.

(3) Pricing of short-term bond: $(1,1,0) \in \mathcal{O}_{\mathcal{K}_{MR_\lambda}}$.

(4) Pricing of risky return: $(1,1,1) \in \mathcal{O}_{\mathcal{K}_{MR_\lambda}}$.

(5) Statistically observable risky and long-horizon bond returns: $(t,0,r) \in \mathcal{O}_{\mathcal{K}_{MR_\lambda}}$ for $(t,r) \in \mathbb{R}^2$.

Figure 2.2 illustrates the observable set $\mathcal{O}_{\mathcal{K}_{MR_\lambda}}$ implied by assumptions (1)-(5). The vertical red plane in (m^T, r) coordinates is generated by observability assumption (5). This assumption follows from the CGF condition:

$$\mathcal{K}_{MR_\lambda}(t,0,r) = \log E[(M^T)^t R_\lambda^r] = \log E[R_\infty^{-t} R_\lambda^r] \,, \tag{2.6.5}$$

which reflects Alvarez and Jermann, 2005 identification $M^T = 1/R_\infty$. The remaining points in the graph, highlighted with purple circles, correspond to assumptions (1)-(4). To generate risky portfolio return R_λ, we consider distinct sets of benchmark assets. Return $R_{\lambda_A} := R_0 + \lambda_A(R_M - R_0)$ (Set A = Mkt + Bond) is the return of a portfolio invested in the short-term zero bond and the aggregate equity market. Return $R_{\lambda_B} := R_0 + \lambda_{B1}(R_1 - R_0) + \lambda_{B2}(R_2 - R_0)$ is the return of a portfolio invested in the short-term zero bond and two size-sorted portfolios with book-to-market ratio in the top 50% quantile of the CRSP universe (Set B = S-G + L-G + Bond).[30] Finally, return $R_{\lambda_C} := R_0 + \sum_{i,j=1}^2 \lambda_{Cij}(R_{ij} - R_0)$ is the return of a portfolio invested in the short-term bond and four double-sorted portfolios with respected to size and book-to-market (Set C = S-G + L-G + S-V + L-V + Bond).[31] We focus on the restrictions

[30] For brevity, S (L) stays for **S**mall (**L**arge) stocks and G (V) for **G**rowth (**V**alue) stocks.

[31] We obtain these stock return data from Kenneth French's website. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

implied by Assumption 1 for the marginal arbitrage-free CGF of pricing kernel vector $M = (M^T, M^P)$ on domain $D = (0,1)^2$. As $D \subset \overline{\mathcal{O}}_{\mathcal{K}_{MR_\lambda}}$, the upper arbitrage-free CGF $\mathcal{K}^U_{MR_\lambda}$ is finite on D and provides a useful observable upper bound for $\mathcal{K}_M$ on this domain. In order to obtain the sharpest upper bound $\mathcal{K}^U_{MR_{\lambda^\star}} = \inf_\lambda \mathcal{K}^U_{MR_\lambda}$ in our tests of the LRR model, we optimize the portfolio composition λ, with respect to weights λ_u corresponding to benchmark investment sets $u = A$, $u = B$ and $u = C$, respectively.

2.6.4 Omnibus Diagnostic Tests of Null Hypothesis $\mathcal{H}_0(\overline{\mathcal{O}}_{\mathcal{K}_{MR_{\lambda^\star}}})$

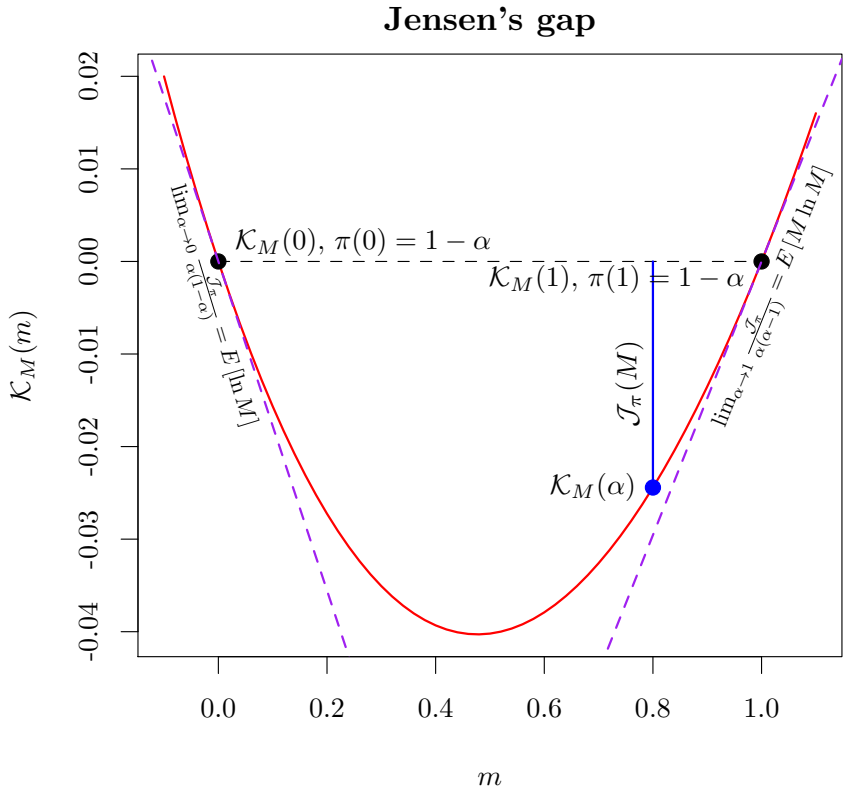
Motivated by the diagnostics testing approach of Section 2.4.6, we first focus on dispersion constraints of Type (1) for the LRR model, over the domain $D \subset \overline{\mathcal{O}}_{\mathcal{K}_{MR_{\lambda^\star}}}$:

$$\mathcal{H}_0(\overline{D}) \; : \; \mathcal{K}^{\mathbb{M}}_M |_D \leq \mathcal{K}^U_{MR_{\lambda^\star}} |_D \; . \tag{2.6.6}$$

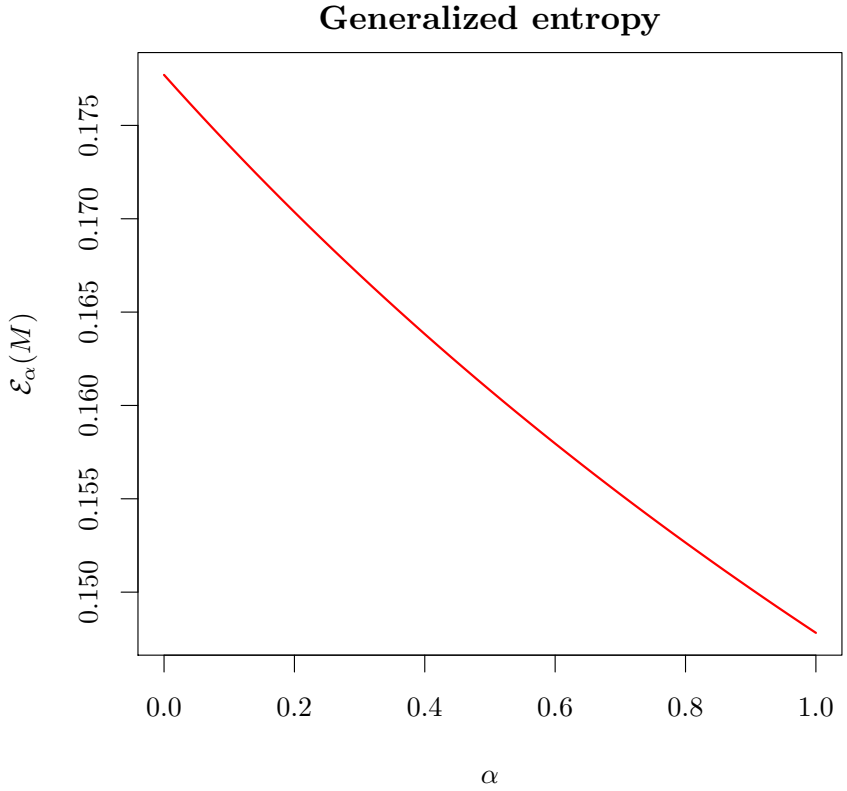
The estimated upper arbitrage free CGF $\hat{\mathcal{K}}^U_{MR_{\hat{\lambda}^\star}} |_D$ for data sets A and C is presented in the middle and the right panel of Figure 2.6, respectively. Similar to the model-implied CGF, the upper CGFs imply a pronounced convexity along the p-axis and a flatter profile along the t-axis, supporting a dominating dispersion of permanent relative to transient pricing kernel components in the data. However, shapes and levels of the upper and model-implied CGFs are also different in a number of cases.

In order to rely on an accurate finite-sample inference on the LRR model, we develop a suitable bootstrap procedure for estimating bootstrap confidence intervals about point estimate $\widehat{\mathcal{K}}^U_{MR_{\hat{\lambda}^\star}}(p,t,0)$ for each $(p,t) \in D$.[32] Conservative bootstrap p-values for the test of null hypothesis (2.6.6) are presented in Figure 2.8a. Using dataset A, we obtain no significant violation of this null hypothesis over the vast majority of domain D. In constrast, using datasets B and C, we progressively obtain wider regions of rejection of null hypothesis (2.6.6), also in the interior of domain D, at standard significance levels. While the violations highlighted by investment sets B and C indicate a possible misspecification of the joint CGF of (M^T, M^P) in the LRR model, it is useful to recall that null hypothesis (2.6.6) is not robust with respect to the model-implied scale of $M^T = 1/R_\infty$. Therefore, a violation of this hypothesis is not directly interpretable as a (scale-invariant) dispersion violation or as an observable scale discrepancy between the model-implied and the arbitrage-free CGF. Similarly, the non rejection of null hypothesis (2.6.6) using investment set A could be the consequence of a low power of the omnibus test in the joint presence of a dispersion violation and an observable scaling discrepancy. Consistently with the concepts developed in Section 2.4, we adress the separation of model violations due to inappropriate dispersion from those due to a scaling discrepancy in the next sections.

[32] Details on this bootstrap procedure are provided in the Supplemental Appendix, available from the authors upon request.

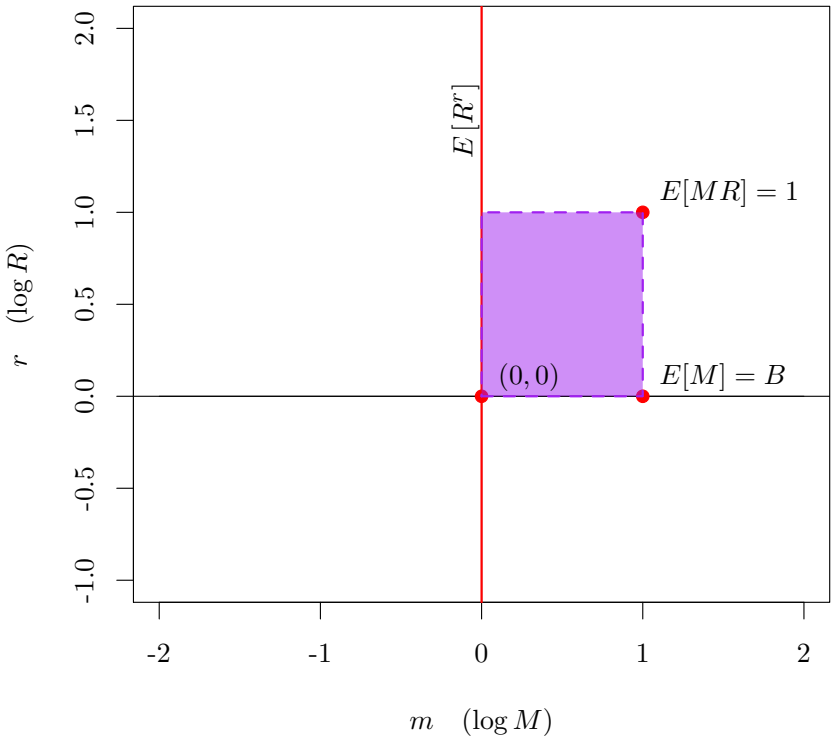


(A) Illustration of Jensen's gap $\mathcal{J}_\pi$ for stochastic discount factor M. $\mathcal{J}_\pi$ at α is calculated with prior π such that $\pi(0) = 1-\alpha$ and $\pi(1) = \alpha$. The vertical blue line at $\alpha = 0.8$ is the gap. At $\alpha = 0$ ($\alpha = 1$) we take limits of the dispersion measure in panel 2.3b obtained by standardising $\mathcal{J}_\pi$ by $\alpha(1-\alpha)$ and illustrate them as the negative slope (slope) of the CGF at 0 (1).

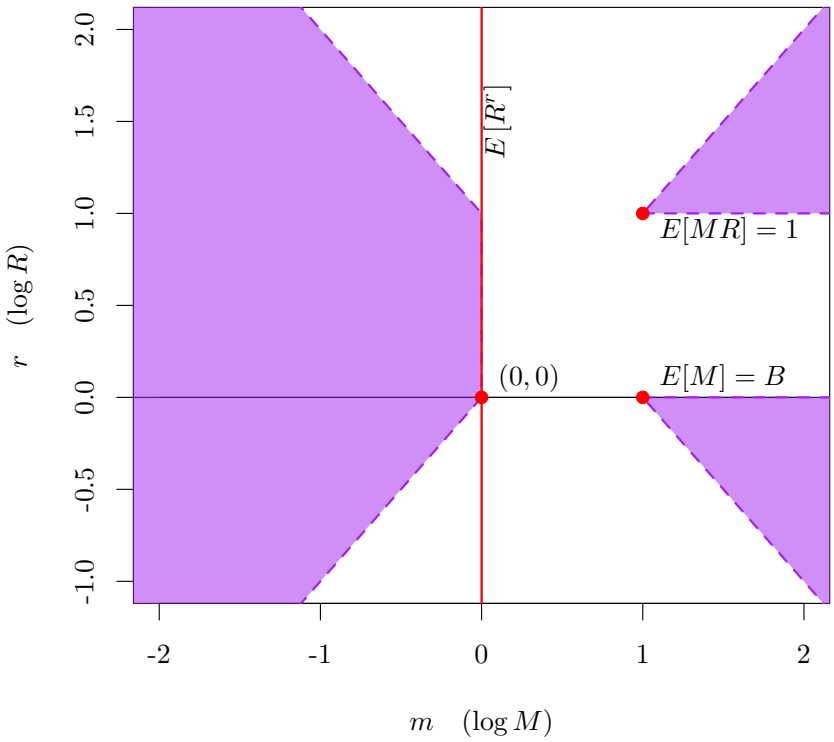


(B) Dispersion $\mathcal{D}_{\pi_\alpha}(M) = \mathcal{J}_{\pi_\alpha}/\alpha(1-\alpha)$ calculated according to equation (2.3.8).

FIGURE 2.3: Illustration of Jensen's gap and generalized entropy of the SDF in equation (2.3.8). The SDF M is centered so that $E[M] = 1$. The standard deviation of M is set to 0.65.



(A) Convex hull $\overline{\mathcal{O}}$ in (M, R) space generated by the observable set $\{(0, r) : 0 \leq r \leq 1\} \cup \{(1, 0), (1, 1)\}$. $\mathcal{K}^U$ is finite on this region.



(B) Set $\underline{\mathcal{O}}$ in (M, R) space generated by the observable set $\{(0, r) : 0 \leq r \leq 1\} \cup \{(1, 0), (1, 1)\}$. $\mathcal{K}^L$ is finite on this region.

FIGURE 2.4: Illustration of regions with finite $\mathcal{K}^U$ and $\mathcal{K}^L$ in a setting with a univariate pricing kernel M and a single priced return R, with an observed bond price B.

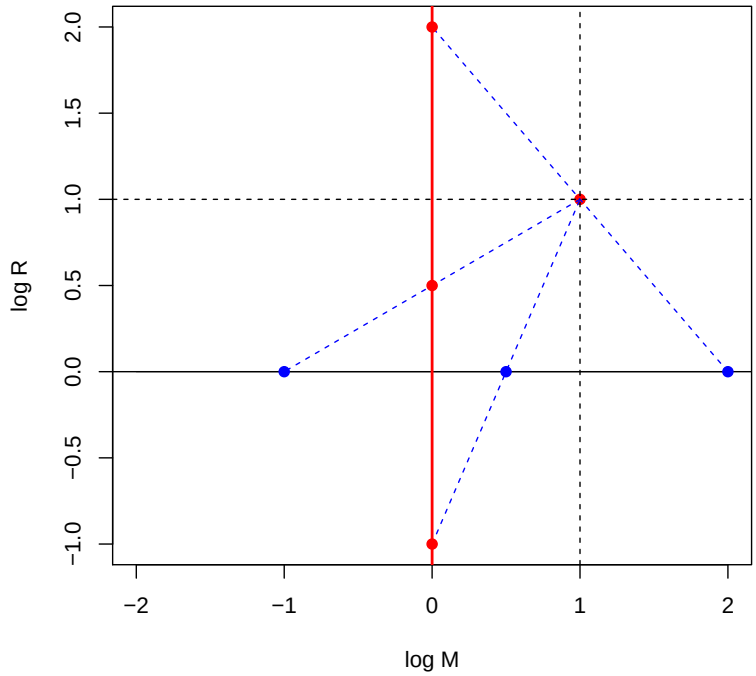


FIGURE 2.5: Formation of dispersion constraints of Type (1) and (2) in a setting with $d_1 = d_2 = 1$. A Type (1) constraint is used to bound the CGF value from above at unobservable point $(1/2, 0)$ with the use of values at observable points $(1, 1)$ and $(0, -1)$. Type (2) constraints are used to bound the value of the CGF from below at unobservable points $(-2, 0)$ and $(-1, 0)$, with priors having support on points $\{(1, 1), (0, 2)\}$ and $\{(1, 1), (0, 1/2)\}$, respectively.

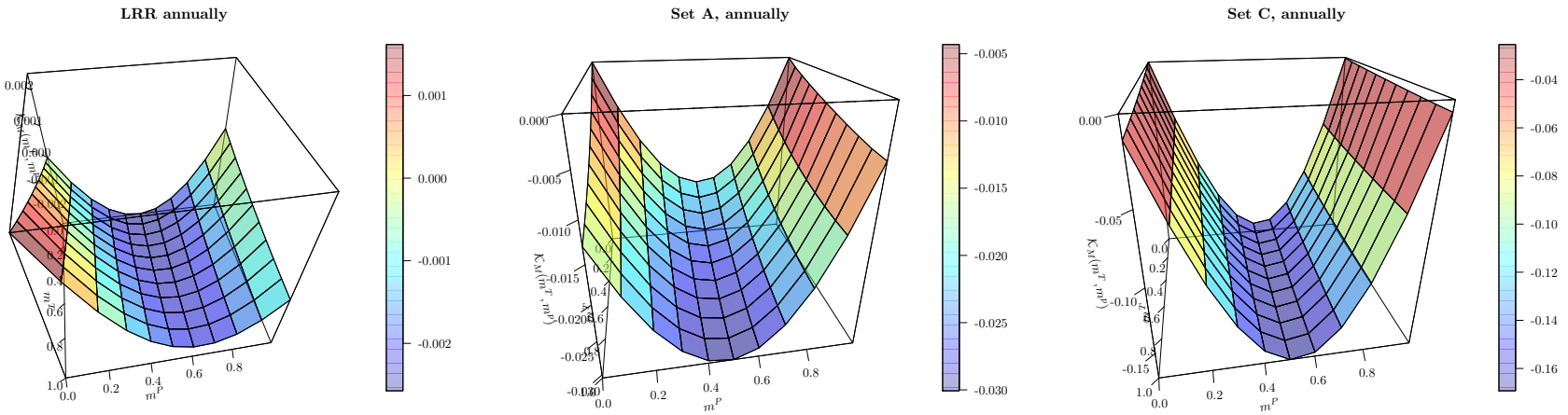


FIGURE 2.6: Model-implied marginal CGF of the SDF components (Bansal, Kiku, and Yaron, 2016, leftmost panels), and estimates of $\mathcal{K}^U_{M^T M^P}(m^T, m^P)$, $(m^T, m^P) \in (0,1)\times(0,1)$, obtained as in Proposition 4. The middle panel presents results where the portfolio of assets contains the market index, the single-period bond and the long-term bond (data set A). The rightmost panel presents results where the portfolio of assets contains, additionally, size- and value- sorted Fama-French portfolios (data set C). Data set description is available in Section 2.6.3.

2.6.5 Diagnostic Test of a Scaling Discrepancy in $M^T = 1/R_\infty$

Due to the normalization of the permanent pricing kernel component in the LRR model, a potential model-implied scaling discrepancy can only arise from an inappropriate scale of $M^T = 1/R_\infty$. Empirically, a scaling discrepancy in M^T can follow from the fact that in the data there is no exact analogue of the long maturity bond return R_∞. For instance, a natural empirical proxy for R_∞ is the return R_{LT} of a bond with the highest observed maturity. Thus, a discrepancy between scales $E[1/R_\infty]$ and $E[1/R_{LT}]$ can be motivated by a limitation of proxy $1/R_{LT}$ for measuring the scale of $1/R_\infty$, rather than by a broader empirical violation of the LRR model.

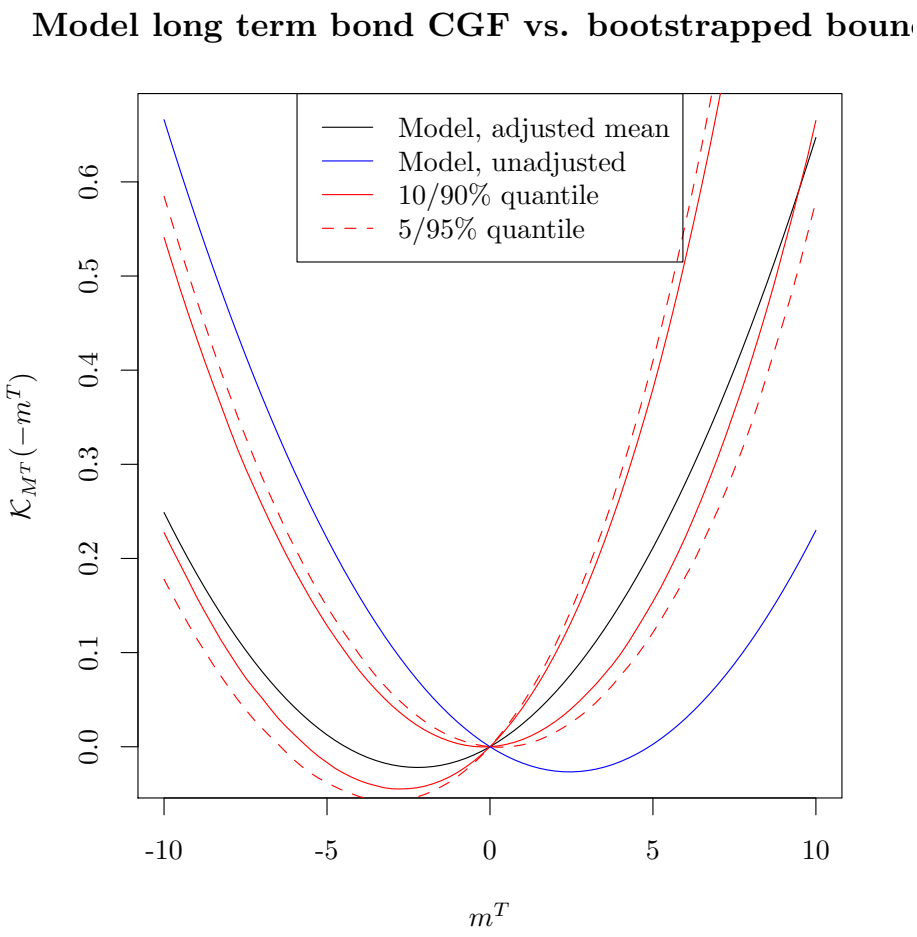
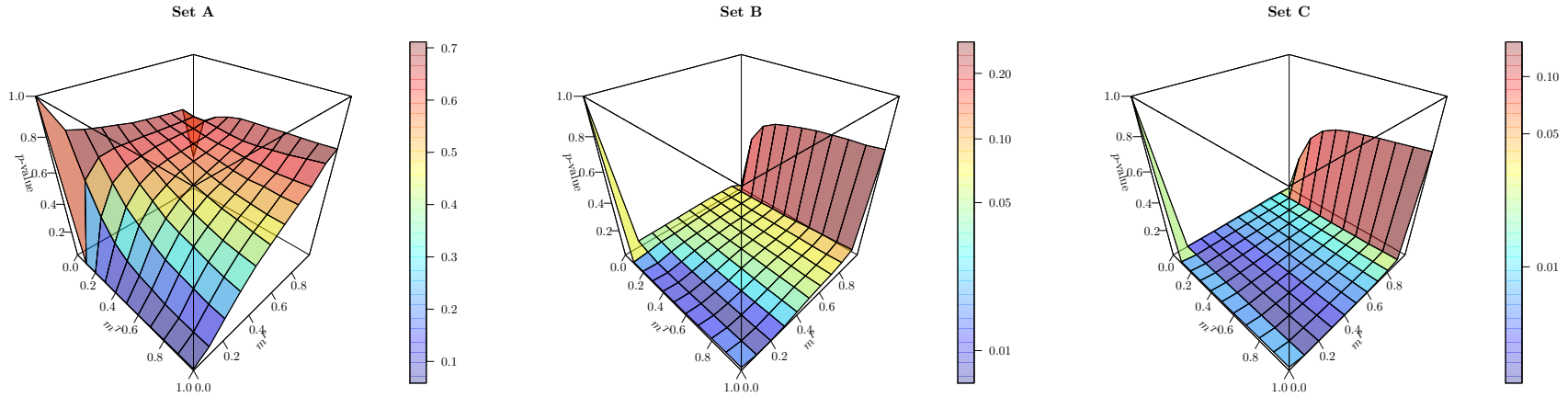
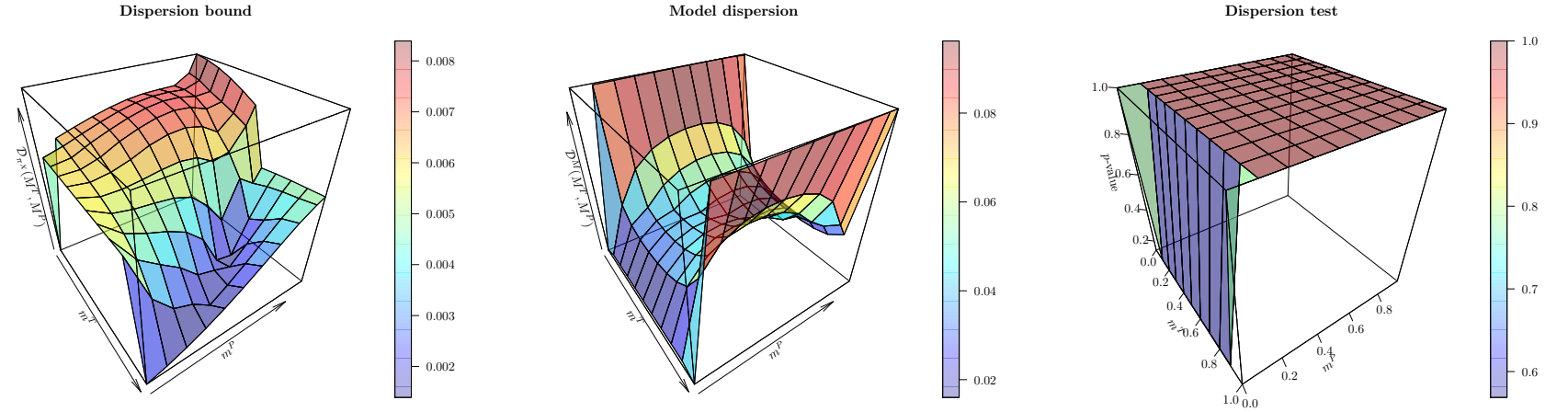


FIGURE 2.7: Model-based and estimated CGF confidence bounds of the transitory component of the stochastic discount factor, with M^T estimated with real returns on long-maturity bonds. Bootstrapped confidence bounds are denoted by red solid and dashed lines. The LRR model-implied CGF is given by the blue line. The black line depicts the LRR model implied CGF once the mean effective return is matched to the data mean (through a constant translation of logarithmic returns).



(A) The omnibus test (see Section 2.6.4): p-values, for $(m^T, m^P) \in (0,1) \times (0,1)$, of the test whether the model-implied CGF evaluated at (m^T, m^P) attains lower values than the upper bound based on Proposition 4. The color scale on the four rightmost panels is logarithmic. Data set description is available in Section 2.6.3.



(B) Dispersion test in the marginal SDF space (see Section 2.6.7): p-values, for $(m^T, m^P) \in (0,1) \times (0,1)$, of the test of the lower dispersion bound (2.6.10) using only observable SDF component information. Annual frequency. Data set description is available in Section 2.6.3.

FIGURE 2.8: Omnibus and marginal SDF space dispersion tests.

We can address the properties of $1/R_{LT}$ as an empirical proxy for $1/R_\infty$, by comparing in Figure 2.7 their marginal CGFs. We find that the marginal CGF of $1/R_\infty$ in the LRR model (plotted in blue) is outside the 95% pointwise bootstrap confidence interval around the empirical marginal CGF of $1/R_{LT}$ (plotted in black). However, a rescaled transient component $\widetilde{M}_T := K/R_\infty$ ($K > 0$) such that $E^{\mathbb{M}}[\widetilde{M}^T] = E[1/R_{LT}]$ produces a marginal model-implied CGF well-inside the bootstrap confidence intervals.[33] Therefore, while a violation of Type (1) due to scaling can be explained by the limitations of proxy $1/R_{LT}$ for measuring the scale of $1/R_\infty$, a violation due to inappropriate dispersion cannot be explained by the limitations of proxy $1/R_{LT}$ for measuring the dispersion of $1/R_\infty$. Indeed, we can directly measure the dispersion of $M_T = 1/R_\infty$, e.g., using the generalized entropy $\mathcal{E}_\alpha(M^T)$ in Section 2.3.2, based on a Bernoulli prior π_α such that $\pi_\alpha(1,0,0) = \alpha \in (0,1)$ and $\pi_\alpha(0,0,0) = 1-\alpha$:

$$\mathcal{D}_{\pi_\alpha}(M^T) = \frac{E_{\pi_\alpha}\left[\mathcal{K}_{M^T}(t)\right] - \mathcal{K}_{M^T}\left(E_{\pi_\alpha}[t]\right)}{\alpha(1-\alpha)} = \frac{\alpha\mathcal{K}_{M^T}(1) - \mathcal{K}_{M^T}(\alpha)}{\alpha(1-\alpha)} = \mathcal{E}_\alpha(M^T)\,, \tag{2.6.7}$$

which is observable under Assumption 1 (5). Figure 2.9 collects the results of a diagnostics test of null hypothesis $\mathcal{H}_0 : \mathcal{E}_\alpha(1/R_\infty) = \mathcal{E}_\alpha(1/R_{LT})$, for parameters $\alpha \in (0,1)$. We find that the model-implied dispersion of $1/R_\infty$ is well inside the 95% bootstrap confidence interval of the estimated dispersion $\widehat{\mathcal{E}}_\alpha(1/R_{LT})$, confirming that the dispersion of $1/R_\infty$ well reproduces the empirical dispersion of $1/R_{LT}$.

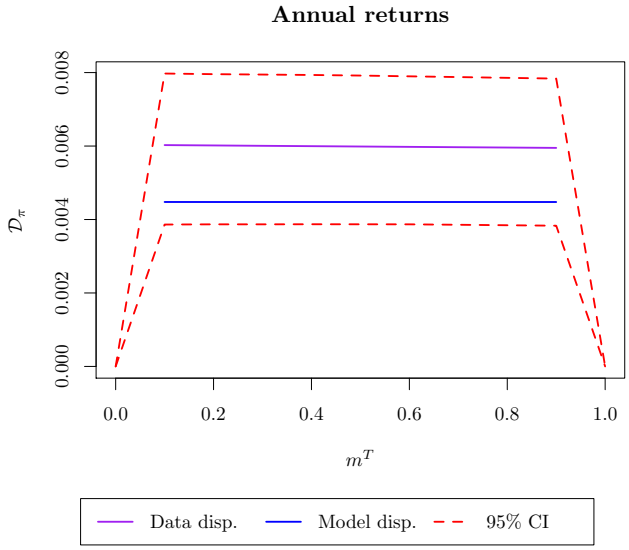


FIGURE 2.9: Observable dispersion of the transitory part of the SDF (2.6.7), calculated under the assumption that real returns on nominal long-maturity bonds are a proxy for real returns on infinite-maturity real bonds, which in turn are the inverse of the transitory component of the SDF, i.e. $R_\infty = \left(M^T\right)^{-1}$.

[33] A rescaling from M^T to $\widetilde{M}^T$ is equivalent to modifying the model-implied CGF by a linear function such that $\mathcal{K}^{\mathbb{M}}_{M^T}(1) = \mathcal{K}_{\widetilde{M}^T}(1)$.

2.6.6 Diagnostics Tests of Observable Excess Dispersion

From Definition 8, an excess dispersion $\Delta\mathcal{J}_{\pi_1,\pi_2}$ is observable when two priors with identical mean and support in $\mathcal{O}_{\mathcal{K}_{MR}}$ exist. In the LRR model, we obtain a family of observable excess dispersions for bivariate vector $M = (M^T, M^P)$, using priors π_1, π_2 such that π_1 has support in $\{(1,1), ((t-p)/(1-p), 0)\}$, π_2 has support in $\{(0,1), (t/(1-p), 0)\}$, and $\pi_1(1,1,0) = \pi_2(0,1,0) = p \in (0,1)$. The resulting observable excess dispersion is:

$$\Delta\mathcal{J}_{\pi_1,\pi_2} = p \log B - (1-p) \log \left(E\left[R_\infty^{-\frac{t}{1-p}}\right] / \, E\left[R_\infty^{-\frac{t-p}{1-p}}\right] \right) , \tag{2.6.8}$$

where we used Assumption 1 (5) to write $M^T = 1/R_\infty$. Similar to the observable covariance measure $\text{cov}(M^T, M^P) = B - E[1/R_\infty]$ in Bakshi and Chabi-Yo (2014), excess dispersion (2.6.8) is not in general robust with respect to the scaling of $1/R_\infty$. However, for $t = p = 1/2$ we obtain the scale invariant observable excess dispersion in Example 2:

$$\Delta\mathcal{J}_{\pi_1,\pi_2}(R_\infty) := \frac{1}{2} \log \left(\text{cov}(R_\infty / E[R_\infty], M^P) + 1 \right) . \tag{2.6.9}$$

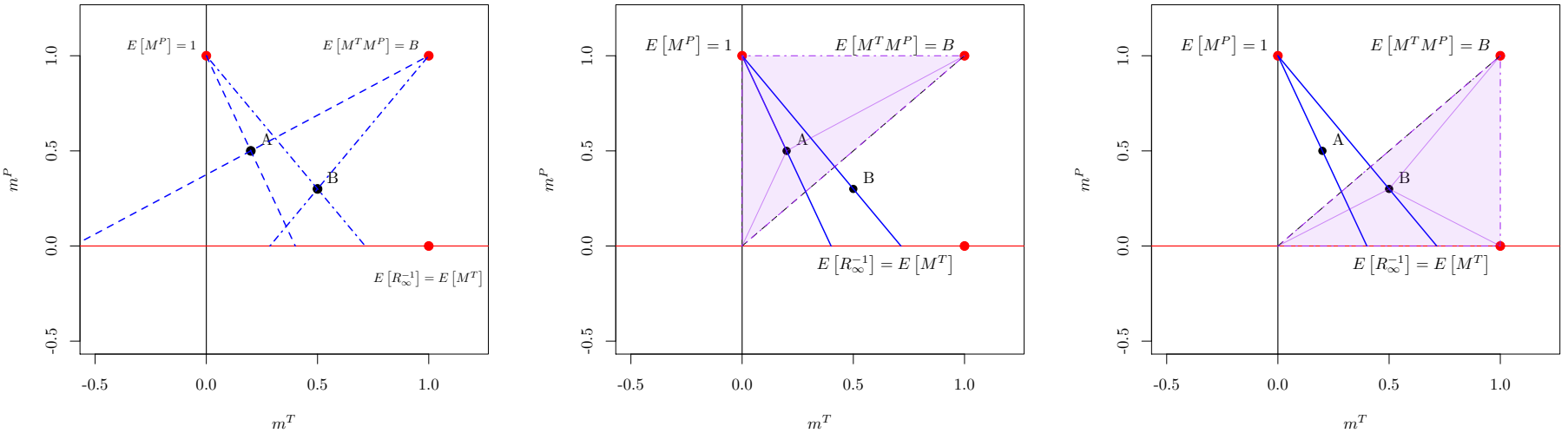
Table A.7 presents the results of a direct test of null hypothesis $\mathcal{H}_0(\Delta\mathcal{J}) : \Delta\mathcal{J}_{\pi_1,\pi_2}(R_{LT}) = \Delta\mathcal{J}^{\mathbb{M}}_{\pi_1,\pi_2}(R_\infty)$ based on excess dispersion (2.6.9). We find that the model-implied excess dispersion is negative, while its point estimate $\widehat{\Delta\mathcal{J}}_{\pi_1,\pi_2}(R_{LT})$ in the data is positive. The difference is statistically significant, as the model-implied excess dispersions does not fit within the 95% bootstrap confidence interval about $\widehat{\Delta\mathcal{J}}_{\pi_1,\pi_2}(R_{LT})$. As excess dispersion (2.6.9) is robust with respect to the scale of M_T, such a violation is independent of the model's ability to fit average long-maturity bond returns and reflects a deeper failure of the LRR model in explaining the unconditional slope of the yield curve.

2.6.7 Diagnostics Tests of Marginal Lower Dispersion Bounds

As illustrated by Figure 2.10, for any point $(t_0, p_0) \in D := (0,1)^2$ we can obtain multiple dispersion constraints of Type (1), using distinct priors π with support in $\mathcal{O}_{\mathcal{K}_M}$ and mean $E_\pi[(t,p)] = (t_0, p_0)$, which exclusively use information from the permanent and transient pricing kernel components. Given set $\Pi(t_0, p_0)$ of such priors and $\pi_0 \in \Pi(t_0, p_0)$, following observable lower dispersion bound emerges:

$$\begin{aligned} \mathcal{D}_{\pi_0}(M) &= \frac{E_{\pi_0}[\mathcal{K}_M(t,p)] - \mathcal{K}(E_{\pi_0}[(t,p)])}{tr(Var_{\pi_0}(t,p))} \\ &\geq \frac{E_{\pi_0}[\mathcal{K}_M(t,p)] - \inf_{\pi \in \Pi(t_0,p_0)} E_\pi[\mathcal{K}_M(t,p)]}{tr(Var_{\pi_0}(t,p))} =: \mathcal{D}^L_{\pi_0,\Pi(t_0,p_0)}(M) \end{aligned} \tag{2.6.10}$$

where the last strict inequality holds if the infimum is not attained in π_0.



(A) Bounds in marginal CGF space – two-point based.

(B) Bounds in marginal CGF space – above the diagonal.

(C) Bounds in marginal CGF space – below the diagonal.

FIGURE 2.10: Dispersion bounds in the marginal CGF space. Red points and lines depict the observable set $\mathcal{O}_{\mathcal{K}_{MR}}$. Black points belong to the convex span of the observable set, $\overline{\mathcal{O}}_{\mathcal{K}_{MR}}$. In 2.10a the value of the CGF in each point is bounded in two ways: by taking the CGF value at $(1,1)$ (the log-bond price) or at $(0,1)$ (the martingale restriction), respectively, and the corresponding points on the m^T axis. In 2.10c and 2.10b the purple triangles provide two more ways of constructing Type (1) dispersion bounds. In order to construct $\mathcal{K}_{MR}^U$ one has to pick the lowest available bound value.

Consistently with the graphical description in Figure 2.10, we construct set $\Pi(t_0, p_0)$ by including the following priors. First, we include prior π_1 with support in $\{(1,1), ((t_0 - p_0)/(1-p_0), 0)\}$ and prior π_2 with support in $\{(0,1), (t_0/(1-p_0), 0)\}$ such that $\pi_1(1,1) = \pi_1(0,1) = p_0$ (see Figure 2.10a). Second, for any $0 < t_0 \leq p_0 < 1$ (any $1 > t_0 > p_0 > 1$) we include priors π_3 with support in $\{(0,0),(0,1),(1,1)\}$ (in $\{(0,0),(1,0),(1,1)\}$) such that $\pi_3(0,0) = (1-t_0)$ and $\pi_3(0,1) = t_0 - p_0$ ($\pi_1(0,0) = (1-t_0)$ and $\pi_1(1,0) = t_0 - p_0$); see Figure 2.10c (Figure 2.10b).

The left and the middle panels of Figure 2.8b plot for any $(t_0, p_0) \in D$ the estimated lower bound $\widehat{\mathcal{D}}^L_{\pi_0,\Pi(t_0,p_0)}(M)$ and the model-implied dispersion $\mathcal{D}^{\mathbb{M}}_{\pi_0}(M)$, respectively, where prior $\pi_0 = \arg\sup_{\pi\in\Pi(t_0,p_0)} E_\pi[\mathcal{K}_M(t,p)]$ is selected to ensure a non trivial lower dispersion bound (2.6.10). We find that typically the estimated lower bound is much lower than the model-implied dispersion, with bootstrap p-values for the test of null hypothesis $\mathcal{D}^{\mathbb{M}}_{\pi_0}(M) \geq \mathcal{D}^L_{\pi_0,\Pi(t_0,p_0)}(M)$, in the right panel of Figure 2.8b, which never produce a rejection even at large significance levels. This evidence shows that the information in the marginal distribution of pricing kernel components, which is generated by the return of short-term and long-maturity bonds, is insufficient to reject the CGF specification of the LRR model in unobservable parts of its domain.

2.6.8 Diagnostics Tests of Joint Lower Dispersion Bounds

Marginal lower dispersion bound (2.6.10) can be substantially improved, based on the information generated by traded portfolio return R_λ, giving rise to joint lower dispersion bounds. Indeed, for $(t_0, p_0) \in D$ we can obtain additional multiple dispersion constraints of Type (1), using the joint arbitrage-free CGF of pricing kernels and returns and a prior π with support in $\mathcal{O}_{\mathcal{K}_{MR_\lambda}}$ such that $(t_0, p_0, 0) = E_\pi[(m, r)]$. For instance, using prior π_1 in the previous section, we obtain:

$$\begin{aligned}\mathcal{D}_{\pi_1}(M) &= \frac{E_{\pi_1}[\mathcal{K}_M(t,p)] - \mathcal{K}_M(t_0,p_0)}{tr(Var_{\pi_1}(t,p))} \\ &\geq \frac{E_{\pi_1}[\mathcal{K}_M(t,p)] - \mathcal{K}^U_{MR_\lambda}(t_0,p_0,0)}{tr(Var_{\pi_1}(t,p))} =: \mathcal{D}^L_{\pi_1}(M,R_\lambda) > 0\,, \quad (2.6.11)\end{aligned}$$

where $\mathcal{D}^L_{\pi_1}(M,R_\lambda)$ is observable empirically. By construction, $\mathcal{K}^U_{MR_\lambda}(t_0,p_0,0)$ is finite and has the following explicit expression from Proposition 5: [34]

$$\begin{aligned}\mathcal{K}^U_{MR_\lambda}(t_0,p_0,0) &= (1-p_0)\mathcal{K}_{M^TR_\lambda}\left(\frac{t_0-p_0}{1-p_0}, -\frac{p_0}{1-p_0}\right) \\ &= (1-p_0)\mathcal{K}_{R_\infty R_\lambda}\left(\frac{p_0-t_0}{1-p_0}, -\frac{p_0}{1-p_0}\right)\,. \quad (2.6.12)\end{aligned}$$

[34] Finiteness of $\mathcal{K}^U_{MR_\lambda}(t_0,p_0,0)$ follows from the fact that $(t_0,p_0,0)$ is in the convex hull of set $\{(t,0,r) : t,r \in \mathbb{R}\} \cup \{(1,1,1),(1,1,0)\} \subset \mathcal{O}_{\mathcal{K}_{MR_\lambda}}$. Figure 2.2 illustrates the observable points in set $\mathcal{O}_{\mathcal{K}_{MR_\lambda}}$.

By optimizing portfolio weight vector λ as in Proposition 4, we obtain the sharpest dispersion bound:

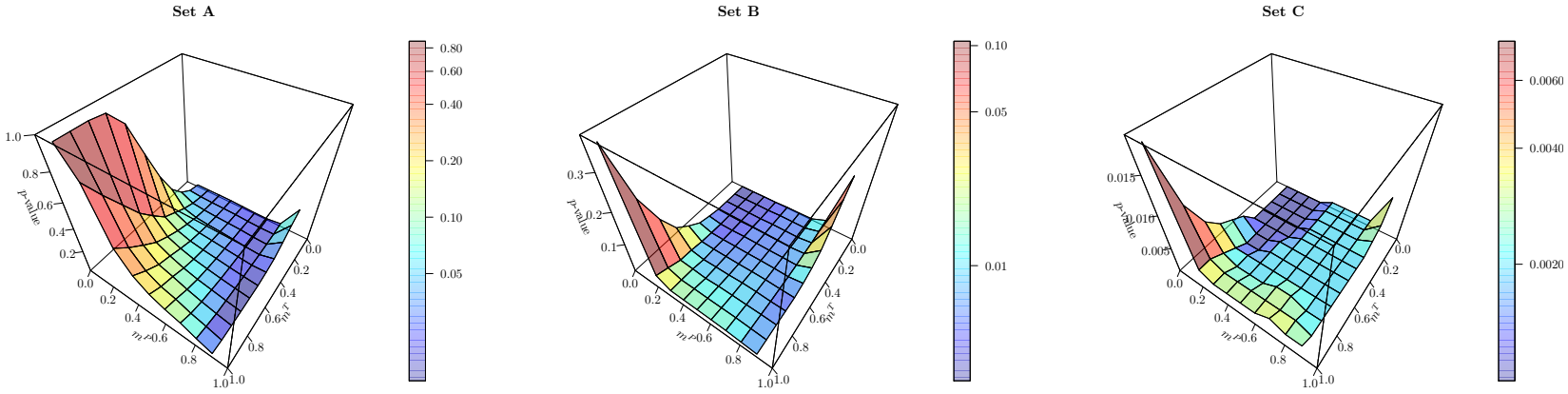
$$\mathcal{D}_{\pi_1}(M) \geq \mathcal{D}^L_{\pi_1}(M, R_{\lambda^\star}) \,, \tag{2.6.13}$$

where $\lambda^\star = \arg\inf_\lambda \mathcal{K}^U_{MR_\lambda}(t_0, p_0, 0)$. This bound is also robust with respect to the scale of R_∞, whenever portfolio return $R_{\lambda^\star}$ does not depend on a position in the long-maturity bond, giving rise to the scale-independent null hypothesis:[35]

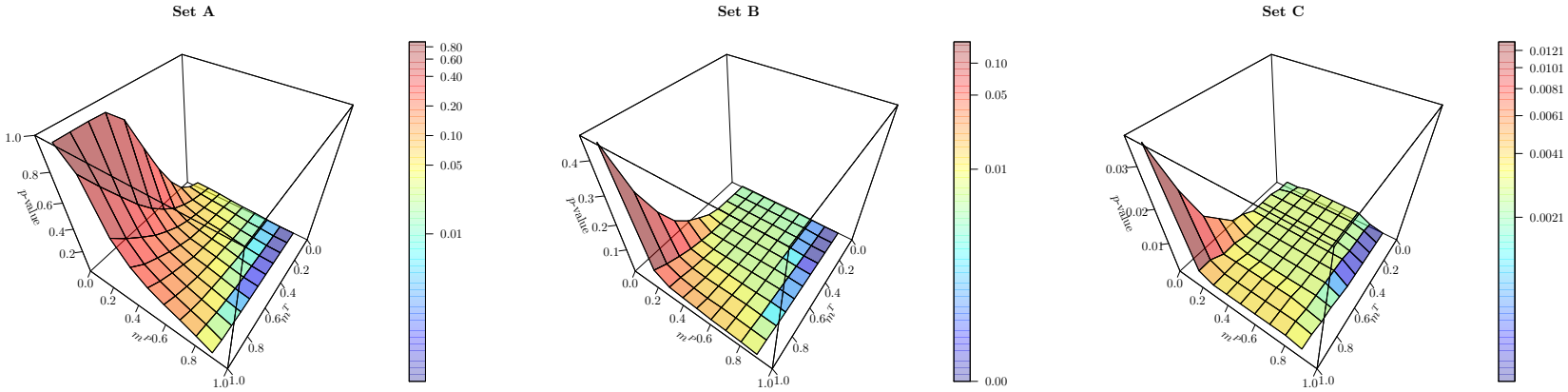
$$\mathcal{H}_0 \,:\, \mathcal{D}^{\mathbb{M}}_{\pi_1}(M) \geq \mathcal{D}^L_{\pi_1}(M, R^\star_\lambda) \,. \tag{2.6.14}$$

Figure 2.11 summarizes the test results of null hypothesis (2.6.14), using investment sets A, B and C and for $(t_0, p_0) \in D$. For each dataset, we obtain a substantial variability of the test $p-$values over domain D. The test results are especially interesting in the left panel of Figure 2.11a, which corresponds to the very simple investment set A, consisting of the short-term risk-free bond and the aggregate equity market. In this panel, we obtain a large number of rejections of the lower dispersion bound (2.6.14) at standard significance levels below 5%, basically for all points above the main diagonal in (t, p) coordinates. In this way, we reject the LRR model by using portfolios of asset returns that were used to estimate the model and by additionally incorporating only the information generated by the long-maturity bond return, in order to construct an observable proxy of the transient pricing kernel component. Note that a multivariate dispersion test, able to exploit the decomposition of the pricing kernel into transient and persistent components, is important in order to uncover these violations. Indeed, the p-value of the dispersion test for $p, t \to 0$, which correponds to a univariate test of the standard entropy bound (2.5.3), does not imply a model rejection at the 5% significance level.

[35] Scale invariance of the bound follows from the fact that $\mathcal{K}_{R_\infty}\left(-\frac{t-p}{1-p}\right) - \mathcal{K}_{R_\infty R}\left(-\frac{t-p}{1-p}, -\frac{p}{1-p}\right)$ does not depend on the scale of R_∞. Including the long-maturity bond in the investment set makes the lower dispersion bound sharper, but it obscures whether a potential dispersion violation is due to an inappropriate convexity or an inappropriate scaling of the model-implied CGF.



(A) Dispersion test in (M, R) space for annual returns, without R_{LT} in the investor's portfolio; p-values.



(B) Dispersion test in (M, R) space for annual returns, with R_{LT} in the investor's portfolio; p-values.

FIGURE 2.11: Dispersion tests in (M, R) space (see Section 2.6.8): p-values. The null hypothesis (2.6.14) is tested for $D_{\pi_{t,p}}$ as in (2.6.11) for $(t, p) \in (0, 1) \times (0, 1)$. Tests without R_{LT} in the investor's portfolio are immune to the mean level of $\log R_{LT}$ in the data. Data set A considers a portfolio of the value-weighted stock index return and short-term bond. Data set B additionally takes size-sorted Fama-French portfolios. Data set C extends to size- and value- sorted Fama-French portfolios. Data set description is available in Section 2.6.3.

For investment set B, the rejection area expands to almost all of domain D, except regions with either high p and small t or high t and small p. A further expansion of the rejection region to virtually all of D arises for investment set C, when book-to-market-sorted returns are taken into account. This finding highlights a difficulty of the LRR model in explaining the dispersion properties of growth and value returns, which complements the model's ability to produce a "cross-section" of value premia close to the extant CAPM, as documented by Bansal, Kiku, and Yaron, 2016. Finally, when return $R^\star_\lambda$ in null hypothesis (2.6.13) depends on the return of the long-maturity bond, we get in Figure 2.11b even stronger violations for all investment sets, which additionally reflect the weak properties of proxy $1/R_{LT}$ for measuring the scale of $1/R_\infty$, shown in Section 2.6.5.

2.7 Conclusion

We introduce a general theory of arbitrage-free dispersion (AFD) that characterizes the testable properties of multivariate asset pricing models. We measure multivariate dispersion by a family of Jensen's gaps that directly reflect the convexity properties of the arbitrage-free cumulant generating function (CGF) of pricing kernels and returns. We show that the observable asset pricing restrictions and the statistical information on asset returns produce tight constraints on the arbitrage-free CGF and its AFD, which are helful to test multivariate pricing kernel specifications with a unifying approach. While our approach naturally incorporates existing AFD constrains in the literature based on univariate pricing kernel bounds, we show that it naturally extends to general multivariate pricing kernel specifications, incorporating, e.g., transient and persistent pricing kernel components, domestic and foreign state prices or horizon dependence.[36] For general multivariate specifications, we systematically develop a wide family of testable AFD constraints, for which we derive closed-form expressions and optimality properties in a number of concrete model settings. Using a recent estimation of the Bansal and Yaron, 2004 model in the literature, we empirically test the multivariate AFD properties of Long Run Risk models, focusing on the joint model-implied distribution of permanent and transient pricing kernel components. We find that while the arbitrage-free and the model-implied CGF both imply a dominating degree of dispersion associated with the permanent component, the joint model-implied dispersion implies a counterfactual dependence of short vs. long-maturity bond returns and is insufficient for pricing the return of a simple portfolio of short-term riskless bonds and market equity. Such dispersion violations are robust with respect to the quality of empirical proxies for long maturity bond returns and are sharper when pricing the return of optimal portfolios invested in double-sorted size and value stocks.

[36] Univariate pricing kernel bounds that are embedded in our AFD approach include Bansal and Lehmann, 1997, Alvarez and Jermann, 2005, Bakshi and Chabi-Yo, 2012, Liu, 2015, Bakshi and Chabi-Yo, 2014 and Backus, Chernov, and Zin, 2014, among others.

3 State recovery from option data through variation swap rates in the presence of unspanned skewness

3.1 Introduction

In this paper I show how certain established realized variation swaps can be interpreted in the context of the slope of the forward-neutral cumulant generating function of log-returns. I discuss the information content of both the fixed, and the floating legs of the contracts. These swap rates succinctly summarize information available in the implied volatility surface. I leverage the fact that in affine jump-diffusion models the swap rates are themselves affine in the state variables to demonstrate (in a simulated setting) that a reduced-dimension filter based on selected swap rates can outperform a non-linear filter based on the whole option panel in an AJD option pricing model. In what follows, I refer to this set of swaps as *affine contracts*. The recommended contract selection – short-maturity skew swaps augmented with longer-maturity variance swaps – is especially beneficial in settings with unspanned skewness, which occurs in AJDs when stochastic factors driving asymmetric jumps do not drive the Brownian volatility.

The importance of an unspanned skewness factor for modeling index option risk dynamics is underlined by the recent literature. Gruber, Tebaldi, and Trojani, 2015 identify the unspanned risk factor with the dynamics of short-term implied volatility skew through a stochastic correlation factor between volatility processes and show that it is particularly important in the description of risk premia. Andersen, Todorov, and Fusari, 2015 find that an extremely $\mathbb{Q}$-persistent, self-exciting tail risk factor is necessary to describe risk premia implied by out of the money put options. Importantly, the factor only drives premia, and is hardly detectable from return (volatility) dynamics. A high-frequency non-parametric analysis of option implied volatilities in Andersen et al., 2015 finds that markedly different factors impact the pricing of low-struck puts as opposed to at-the-money calls: the former has a jump nature, while the latter is a Gaussian process and is associated with volatility. Calvet et al., 2015 link the jump factor to the dynamics of persistent changes in the level of volatility, again decoupling the jump intensity from the evolution of spot volatility. Li and Zinna, 2017 lean upon the term structure of variance swaps to find that the short end

of the volatility risk premium responds strongly to the jump factor whose intensity is, again, independent of Brownian volatility dynamics.

Typically, the stochastic drivers of time-varying return moments and risk premia are not directly observable, and have to be inferred from (option) data. Extracting that information, however, requires considerable modeling and computational effort. The two dominant approaches in the literature rely either on (approximate) maximum likelihood or on non-linear least squares or GMM. They differ importantly in how the latent states are estimated. The former method requires the knowledge of the transition density of the state variables, and a probabilistic description of observation error to then estimate the states in a Bayesian manner from panels of data. Examples can be found in .e.g. Eraker, Johannes, and Polson, 2003; Calvet et al., 2015; Fulop and Li, 2015; Li and Zinna, 2017. The latter treats state variables as parameters to be estimated jointly with structural model parameters, and is exemplified e.g. in Pan, 2002; Bates, 2000; Andersen, Fusari, and Todorov, 2015a; Andersen, Todorov, and Fusari, 2015. In a third approach, Aït-Sahalia, Karaman, and Mancini, 2015 treat a selection of variance swap contracts as free of observation error and directly solve for the states, inverting the linear mapping postulated by the model.

These estimation approaches differ in what data the researchers take as primitives, and can be grouped in two sets. Studies such as Bates, 2000; Pan, 2002; Eraker, Johannes, and Polson, 2003; Calvet et al., 2015 and Andersen, Fusari, and Todorov, 2015a; Andersen, Todorov, and Fusari, 2015 estimate the models on data on the option implied volatility surface and their focus is on risk premia associated with the factors driving return variation. Aït-Sahalia, Karaman, and Mancini, 2015; Fulop and Li, 2015 and Li and Zinna, 2017 turn their focus to variance risk premia, and can thus limit their attention to variance swap data, which do not encode all the information available in the option panel, however offer a dramatic reduction in computational complexity and the benefits of linear filtering. Augmenting the term structure of variance swaps with other affine contracts retains the simple form of the filtering problem and the information content of the complete implied volatility surface. Furthermore, as opposed to individual options, in the case of affine contracts it is straightforward to identify a contract's loadings on the factors. Such sharpness brings along a potentially improved identification of the factors that have a major impact only on the prices of a subset of options – the risks unspanned by volatility.

The paper proceeds with a discussion of the properties and the interpretation of the forward-neutral cumulant generating function of log-returns in Section 3.2, followed by the description and interpretation of CGF slope swaps 3.3. It presents results of simulation studies of various observation equation specifications in filtering in Section 3.4. Section 3.5 concludes.

3.2 The forward-neutral cumulant generating function

The CGF, through no-arbitrage restrictions, has several interesting properties under the forward-neutral probability measure. I show how under these restrictions the derivatives of the CGF can be interpreted as measures of expected return variation.

The cumulant generating function of the log-return, defined by

$$K(p,\tau|\mathcal{I}_t) := \ln \mathbb{E}\left[(F_{t+\tau}/F_t)^p \,|\mathcal{I}_t\right], \tag{3.2.1}$$

is a convex and continuous function of argument p. I also introduce the following notation, which will simplify many upcoming expressions:

$$\phi_p(x) := x^p. \tag{3.2.2}$$

The CGF uniquely identifies the forward-neutral probability distribution. Therefore, if it is possible to evaluate the CGF at arbitrary p, then it is possible to uniquely determine the state pricing measure. In empirical applications, the forward-neutral CGF has to be estimated from the prices of European option contracts, as in Section 3.3.2. For small and large values of p the estimates require the knowledge of prices deep out of the money put and call contracts, respectively. If they are at all available, these prices are subject to high uncertainty through wide bid-ask spreads and potentially infrequent updating. As a consequence, estimates of CGF values for p outside of the $[0,2]$ interval are highly unreliable. This is because estimating the CGF (its slope) from option prices requires weighing the option prices by weights of order K^{p-2} ($\ln K K^{p-2}$). My experience with option data shows that below $p=0$ ($above=2$) the put (call) option weights grow too fast to ensure decent swap rate approximations with the available option quotes. Fortunately, the CGF and especially its first derivative on the smaller interval $[0,1]$ contain rich information about the second and third moments of simple returns.

An example CGF is plotted in Figure 3.1. Under the forward measure $\mathbb{Q}_{t,\tau}$ the time-t expectation of the forward return is $\mathbb{E}^{\mathbb{Q}_{t,\tau}}\left[F_{t+\tau}/F_t|\mathcal{I}_t\right] = 1$. [1] This property translates to $K_t(0,\tau) = K_t(1,\tau) = 0$ and has important consequences for the interpretation for the CGF. For instance, together with convexity and continuity, it implies that there exists a $p^* \in (0,1)$ which is a global minimum. It follows that the values of the CGF on the unit interval are bounded and easier to estimate.

The derivatives of the CGF evaluated at 0 convey information about the cumulants – and through them, moments – of the log-return. As a consequence of the particular features of the forward measure, and of the concavity of the logarithm function, the first derivative of the CGF evaluated at any p is informative of the second or third moment of *simple* returns $F_{t+\tau}/F_t$. To see that, note the form of $K'_t(p,\tau)$ in equation (3.2.3), and its expansion in simple return, around its forward-neutral

[1] In order to de-clutter notation, I write $\mathbb{Q}$ for $\mathbb{Q}_{t,\tau}$, $K_t(p,\tau)$ for $K(p,\tau|\mathcal{I}_t)$, and $\mathbb{E}_t^{\mathbb{M}}\left[X_{t+\tau}\right]$ for $\mathbb{E}^{\mathbb{M}_{t,\tau}}\left[X_{t+\tau}|\mathcal{I}_t\right]$ whenever the contextual understanding does not suffer.

mean, 1, in equation (3.2.4).

$$K'_t(p,\tau) = \mathbb{E}_t^{\mathbb{Q}}\left[\left(\frac{F_{t+\tau}}{F_t}\right)^p \ln\frac{F_{t+\tau}}{F_t}\right] / e^{K_t(p,\tau)} \tag{3.2.3}$$

$$= \frac{\mathbb{E}_t^{\mathbb{Q}}\left[\left(\frac{F_{t+\tau}}{F_t}-1\right) + \left(p-\frac{1}{2}\right)\left(\frac{F_{t+\tau}}{F_t}-1\right)^2 + \left(\frac{p^2}{2}-p+\frac{1}{3}\right)\left(\frac{F_{t+\tau}}{F_t}-1\right)^3 + \ldots\right]}{e^{K_t(p,\tau)}}. \tag{3.2.4}$$

$K'_t(p,\tau)$ above is the slope of $K_t(p,\tau)$ plotted in the right panel of Figure 3.1. Under an arbitrary probability measure, $K'_t(p,\tau)$ would contain a first moment component. Under no-arbitrage restrictions of the probability measure, the first element of the expansion in (3.2.4) drops out, and the leading order of the expansion depends on p. While it is not immediately obvious from (3.2.3) what this quantity measures, the expansion (3.2.4) is rather instructive.

From expansion (3.2.4), the leading term in the slope of the CGF is determined for almost all ps by the second moment of the forward return. In the left panel of Figure 3.1 I plot the slope at three points of particular interest. The CGF's slope at 0 ($K'_t(0,\tau)$), marked with a green line, is the expectation of $\ln F_{t+\tau}/F_t$. Evaluated at 1 ($K'_t(1,\tau)$), marked with a red line, it is the expectation of $F_{t+\tau}/F_t \ln F_{t+\tau}/F_t$. The leading terms in these quantities are, to first order, determined by the forward-neutral variance of simple returns. They differ in how higher moments of the distribution of simple returns impact them. Importantly, to the left (right) of $p = 1/2$ the second moment of simple returns enters with a negative (positive) sign, therefore the steeper the CGF descends (ascends) at 0 (at 1), the higher the variance of simple returns. The slope at $p = 1/2$ ($K'_t(1/2,\tau)$), marked with a brown line, is a particularly interesting case. Indeed, in the expansion (3.2.4), the leading variance term drops out precisely at $p = 1/2$, which implies that the CGF slope at $p = 1/2$ is a measure of skewness of the simple return.

A precise notion of skewness can be assigned to the CGF slope at $p = 1/2$. Indeed, note that $(F_{t+\tau}/F_t)^p / e^{K_t(p,\tau)}$ can be interpreted as a Radon-Nikodym derivative, and (3.2.3) can be interpreted as the expectation of log return under a changed probability measure. The particular feature of this measure, as shown in Schneider and Trojani, 2015a, is that $K'(1/2,\tau) = 0$ if and only if the forward distribution is put-call symmetric (Carr and Lee, 2009). In the case of put-call symmetry, the minimum of the CGF falls at $1/2$. Whenever the minimum falls to the left (right) of $1/2$, the associated distribution is left-skewed (right-skewed).

In summary, in this section I showed how the CGF and its first derivatives are informative of the second and third moments of the forward-neutral distribution. If the forward measure allows no arbitrage, then prices of payoffs can be found by evaluating the relevant forward expectation. A natural consequence of this fact is that there exist trading strategies whose prices are expressed by $K'(p,\tau)$. I describe the strategies and their payoffs in the Section 3.3.

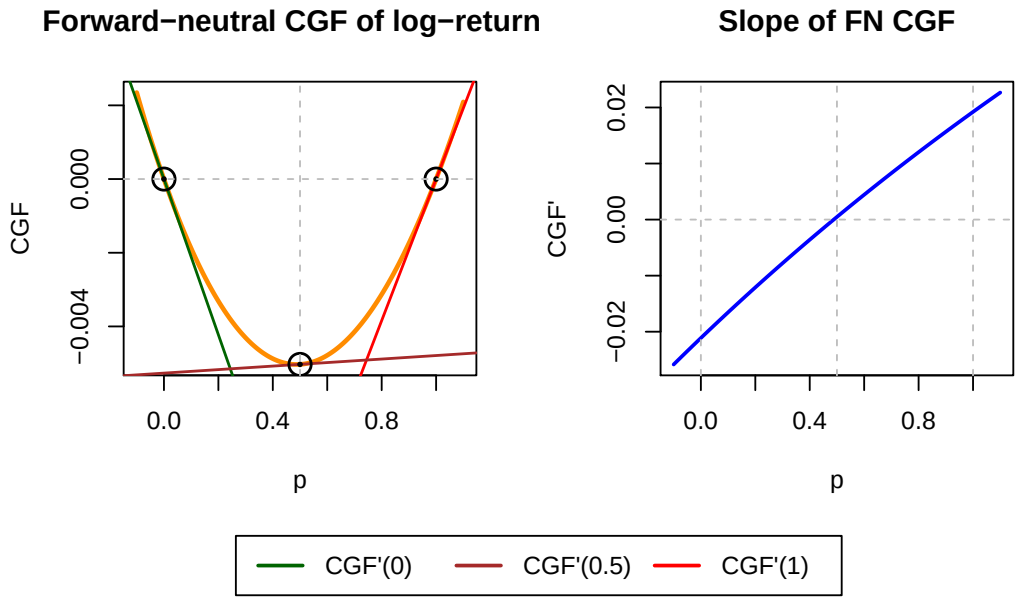


FIGURE 3.1: Example cumulant generating function and its derivative. Derivatives of the CGF at three chosen points are illustrated with tangent lines in the left panel.

3.2.1 CGF in affine jump diffusions

I consider a general affine stock return dynamics, written as:

$$\begin{aligned} \frac{dF_t}{F_{t-}} &= \mu(V_{t-})dt + \Phi(V_{t-})dW_t^S + \int_{\mathbb{R}\setminus\{0\}} (e^{g(x,y)} - 1)\nu_t(dx,dy)dt \\ dV_t &= \kappa(V_{t-})dt + \Sigma(V_{t-})dW_t^V + \int_{\mathbb{R}_+^M} y\nu_t(dx,dy)dt \end{aligned} \tag{3.2.5}$$

with $\mu : \mathbb{R}^M \to \mathbb{R}$, $\kappa : \mathbb{R}^M \to \mathbb{R}^M$, $\Phi'\Phi$, and the diagonal of $\Sigma'\Sigma$ (the off-diagonal elements are 0) affine in V. ν_t is a counting measure whose compensator is affine in V. All components of dW^S are independent, and so are those of dW^V. It is, however, possible to allow for correlation between dW^S and dW^V to model the "leverage effect", that is to reflect the fact that negative returns often accompany icreases in volatility. This notation covers basically all models existing in the literature on standard state spaces.

Duffie, Pan, and Singleton, 2000 and Duffie, Filipovic, and Schachermayer, 2003 give the general form of the joint cumulant generating function of the log-return and variance factors in settings like (3.2.5) and impose necessary technical conditions on the jump measure ν. Borrowing from the intuitions of Section 3.2, I focus on the marginal CGF of the log-return and its derivative. For $p \in \mathbb{R}$, write

$$K_t(p,\tau) = \alpha(p,\tau) + \beta(p,\tau)\cdot v_t \tag{3.2.6}$$

$$K_t'(p,\tau) = \alpha'(p,\tau) + \beta'(p,\tau)\cdot v_t. \tag{3.2.7}$$

The coefficients $\alpha, \alpha', \beta, \beta'$ can be found by solving systems of ordinary differential

equations, as stated in Propositions 1 and 3 in Duffie, Pan, and Singleton, 2000, respectively for the CGF and its derivative.

It is instructive to investigate the sensitivity of the shape of the CGF and its derivative to the parameter values of a simple affine jump diffusion model. From changes in the shape and values of the CGF and its derivative, induced by parameter changes, one can infer (a) which risk-neutral moments of log-returns the parameter changes impact, and (b) how contract prices might reflect higher-order moment information.

The Bates, 1996 specification is the simplest "laboratory" which allows for a rich comparative statics analysis. The model allows for jumps in the underlying (with a constant jump intensity) and a Brownian leverage effect.

$$\begin{aligned} \frac{dF_t}{F_{t-}} &= -\lambda_0 \mathbb{E}[e^x - 1]dt + \sqrt{v_t} dW_t^S + \int_{\mathbb{R}\setminus\{0\}} (e^x - 1)\,\nu(dx)dt \\ dv_t &= \kappa(\eta - v_t)dt + \sigma\sqrt{v_t} dW_t^v \\ dW_t^S dW_t^v = \rho dt, \text{ and} \qquad & \frac{\nu(dx)}{dx} = \frac{\lambda_0}{\sqrt{2\pi}} e^{-\frac{(x-\mu_J)^2}{2\sigma_J^2}} \end{aligned} \tag{3.2.8}$$

I modify the parameters from those contained in the paper so that they better reflect the characteristics of stock returns, rather than foreign currency returns, which served as the estimation sample. I set a negative jump mean and add a substantial leverage effect. The exact values are not of central interest, albeit they are reported in Figure 3.2.

Consider the CGF and its derivative as mappings over the underlying model parameter space, Θ, for fixed state variable value, and for a range of values of p. I adopt the notation $K_t(\theta; p, \tau)$ and $K_t'(\theta; p, \tau)$ for the CGF and its derivative, respectively. Both functions map from Θ to $\mathbb{R}$. A direct graphical representation of both functions for multiple parameter value sets is not informative, as it is hard to visually discern the subtle shape changes induced even by substantial parameter modifications. Therefore, in Figure 3.2 I plot the mappings $\frac{\partial}{\partial\theta_j} K_t(\theta; p, \tau)$ (in orange) and $\frac{\partial}{\partial\theta_j} K_t'(\theta; p, \tau)$ (in blue) for all forward-neutral parameters of the model, while holding the $v_t = 0.02$ constant. These curves, and particularly their values at $p \in \{0, 1/2, 1\}$, illustrate how the forward-neutral moments change with model parameters.

The interpretation of these curves requires some prudence. $\frac{\partial}{\partial\theta_j} K_t'(\theta; p, \tau)$, plotted in blue in Figures 3.2, gives precisely the effect of a parameter change on the risk-neutral moments of returns. Focus on $p \in \{0, 1/2, 1\}$. In the beginning of section 3.2 I showed that at $p = 0$ and $p = 1$ ($p = 1/2$), the leading term in $K_t'(p, \tau)$ is determined by the variance (skew) of returns. Recall, however, that $K_t'(0, \tau) \approx -1/2\mathbb{V}_t[y_{t+\tau}]$, $K_t'(1, \tau) \approx 1/2\mathbb{V}_t[y_{t+\tau}]$, and $K_t'(1/2, \tau) \approx -1/24\mathbb{E}_t[y_{t+\tau} - 1]^3$ for $y_{t,\tau} = F_{t+\tau}/F_t$. It follows that the lower (higher) the blue curve at $p = 0$ ($p = 1$) in Figure 3.2, the more the forward-neutral variance increases with an increase in the parameter. Similarly, the higher the blue curve at $p = 1/2$, the more negative skewness there is in the

model. The parameters $(\lambda_0, \mu_J, \sigma_J, \eta, \kappa)$ have an almost linear impact on the shape of skewness, and the curves pass close to 0 at $p = 1/2$ in Figure 3.2. The values of the blue curves for at 0 and 1 give directly their marginal impact on signed return variance. Thus $(\lambda_0, \mu_J, \sigma_J, \eta, \kappa)$ are the key determinants of return variance in this model. For the remaining parameters, from the magnitudes of the response of the CGF slope I can presuppose whether the perturbation of the parameter changes the third or fourth risk-neutral moment. An increase in the leverage parameter from the initial value of $\rho = -1$ or from $\sigma = 0.38$ has no impact on the expected log-return, and thus on return variance, because the blue curves pass through 0 in Figure 3.2. This implies that the price of the variance swap does not reflect Brownian skewness, and it is also not affected by structural changes in volatility of volatility. The slope of the CGF at $p = 1/2$ decreases in ρ at the starting parameter value, that is an inrease in ρ decreases negative skewness. The two remaining parameters impact the slope of the CGF at $p = 1/2$ via the fourth moment of returns rather than the third. An increase in the volatility-of-volatility parameter σ increases the fourth moment of returns, and the impact is visible in the CGF slope at $p = 1/2$.

Before considering $\frac{\partial}{\partial \theta_j} K_t(\theta; p, \tau)$, note that the CGF values at $p = 0$ and $p = 1$ are anchored at 0 by no-arbitrage. It follows that if the level of the CGF changes at some point inside $p \in [0, 1]$, it must, due to convexity and continuity, entail changes in slope and/or curvature of the CGF at some points. Curvature is directly related to the second derivative of the CGF with respect to the argument p, $K_t''(p, \tau)$, and thus to the second cumulant (variance) of the log-return, which translates to the third or fourth moment of the simple return, depending on p. The shapes of $\frac{\partial}{\partial \theta_j} K_t(\theta; p, \tau)$ in Figure 3.2 tells us whether a change in a parameter value predominantly impacts CGF's curvature or symmetry around $1/2$. Note that $\frac{\partial}{\partial \theta_j} K_t(\theta; p, \tau)$ curves are almost symmetric around $p = 1/2$ for λ_0, μ_J, σ^J, κ and η, as opposed to other parameters. Imagine, for each parameter, adding the orange curves in Figure 3.2 to the CGF plot in the left panel of Figure 3.1. As values of these parameters increase, the CGF becomes flatter (more curved) for μ_J (λ_0, σ_J and η), for κ the sign of the parameter derivative would be opposite were $v_t > \eta$; these parameters predominantly govern return variance. For the remaining parameters, the effect is not symmetric around $p = 1/2$. Again, add the orange curve in the ρ panel of Figure 3.2 to the CGF plot in Figure 3.1. This moves the minimum to the right, in this case decreasing the distance between $p* = \text{argmin} K_t(p, \tau)$ and $1/2$, which is a measure of skewness. In this way, the curve $\frac{\partial}{\partial \theta_j} K_t(\theta; p, \tau)$ shows which parameters predominantly govern return variance, and which govern higher moments, but does not show the magnitude of the change.

The static analysis above, conducted in an example model, shows that, in isolation, the price of a variance swaps conveys little information about leverage or volatility-of-volatility parameters, while the price of a skew swap is not informative of jump intensity, or of the jump variance parameter. By no means is that a

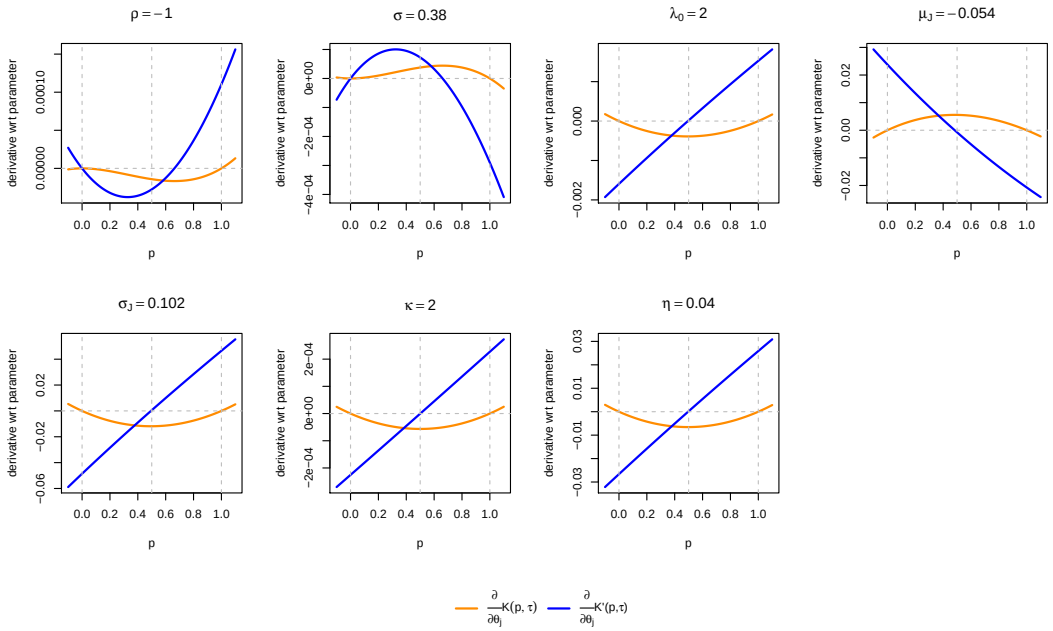


FIGURE 3.2: Comparative statics: sensitivity of cumulant generating function of the log-return, and of its first derivative to changes in model parameters.

general rule, but rather a demonstration of how our intuitions about model parameters translate to the properties of the CGF. The following intuition should be gleaned from this section: in an estimation exercise certain parameters can be better identified from the dynamics of some contracts, than others. In the example model, one would rightly expect that a single observation of the term structure of variance swaps is completely uninformative of the values of ρ and σ. While in an estimation exercise with variance swap and return, these parameters can be identified from, respectively, the dynamics of returns, and the dynamics of variance swaps together with returns, one can strengthen their identification by adding the term structure of skewness swaps to the observation equation, as estimating skewness features from returns is notoriously difficult.

3.3 CGF slope swaps

CGF slope swaps are financial contracts whose price (the fixed leg) is defined through equation (3.2.3), and which have a floating leg defined below. The latter has a non-linear component that arises directly from replicating the term inside the expectation in equation (3.2.3), and a delta-hedging component. Sufficiently frequent delta-hedging allows to interpret the swap as a bet on realized variation of the underlying forward return.

The payoff of a CGF slope swap can be constructed from (3.2.3) by taking the term inside the expectation, delta-hedging it, and adjusting the size of the position

by the numerator:

$$\frac{\overbrace{\left(\frac{F_{t+\tau}}{F_t}\right)^p \ln\frac{F_{t+\tau}}{F_t} - \frac{(F_{t+\tau}-F_t)}{F_t}}^{\text{option position, non-linear payoff}} - \overbrace{\sum_{j=1}^{N}\left(\frac{1}{F_{t_{j-1}}}-\frac{1}{F_t}\right)\left(F_{t_j}-F_{t_{j-1}}\right)}^{\text{frequent delta-hedging}}}{e^{K_t'(p,\tau)}}. \tag{3.3.1}$$

By construction, the forward-neutral expectation of this expression is equal to (3.2.3).

Consider an implementation of the strategy with a single hedging step, i.e. with $t_N = t+\tau$ and $N=1$. For $y := \ln F_{t+\tau}/F_t$ the non-linear component of the payoff in (3.3.1) can be written as $ye^{py} - (e^y - 1)$. A Taylor expansion of the payoff at $y=0$ yields

$$ye^{py} - (e^y - 1) = \sum_{k=1}^{\infty}\left(\frac{p^k}{k!} - \frac{1}{(k+1)!}\right) y^{k+1}. \tag{3.3.2}$$

For all $p \neq 1/2$ the leading order of this expansion is y^2, while for $p = 1/2$ the first element vanishes, and the leading order becomes y^3. Whenever $N > 1$ and forward trading is more frequent, with $y_j = \ln F_{t_j}/F_{t_{j-1}}$, it is possible to write (3.3.1) in terms of sums of $\phi_p(y_j) - 1 - p(e^{y_j} - 1)$ and of $\phi_p'(y_j) - (e^{y_j} - 1)$, weighted by functions of $F_{t_{j-1}}/F_t$. This coincides with the notion of realized divergence, as defined in Schneider and Trojani, 2015a: CGF slope swaps are embedded in the family of power divergence swaps and their higher-order descendants. The total payoff in (3.3.1) has a well-defined limit for ultra-high frequency trading, that is for $\sup|t_j - t_{j-1}| \to 0$, for every $p \in [0,1]$. I give the limits for selected ps below.

The contracts of interest, with $p \in \{0, 1/2, 1\}$ are already known in the literature, albeit to the best of my knowledge they have not been interpreted jointly in the context of the CGF. [2] In what follows, I give formulas of terminal payoffs and their continuous-time limits together with leading references.

$p = 0$: Variance swap (VIX)

The value of the square of the VIX index (CBOE, 2000; Jiang and Tian, 2007) corresponds to $-2\mathbb{E}_t^{\mathbb{Q}}\left[\ln F_{t+\tau}/F_t\right]$, i.e. the $p=0$ slope swap is equivalent to the VIX^2 variance swap, after suitable rescaling and adjustment of the payoff function. The price and payoff of the variance swap are, respectively:

$$\begin{aligned} &-2\mathbb{E}_t^{\mathbb{Q}}\left[\ln F_{t+\tau}/F_t\right] \\ &-2\ln F_{t+\tau}/F_t + \sum_{j=1}^{N}\frac{F_{t_j}-F_{t_{j-1}}}{F_{t_{j-1}}} = -2\sum_{j=1}^{N}\ln\frac{F_{t_j}}{F_{t_{j-1}}} - \frac{F_{t_j}-F_{t_{j-1}}}{F_{t_{j-1}}}. \end{aligned}$$

Typically, this swap rate is used in transactions which exchange it for realized variance for returns, $RV_{t,\tau} = \sum_{j=1}^{N}\left(F_{t_j}/F_{t_{j-1}} - 1\right)^2$. Consider the high frequency limits

[2] Orłowski, Sali, and Trojani, 2015 in fact consider $K_t'(p)$ at $p=0$ and $p=1$ in the context of pricing kernel dispersion bounds.

of the realized variance and the payoff (floating leg) of the $p = 0$ swap as defined above. In the absence of jumps, and after correcting for scale and sign, the two quantities would have exactly the same limit, integrated variance: $\int_t^{t+\tau} \sigma_s^2 ds$. If jumps are present, this holds only approximately. For the details of the approximation, see Jiang and Tian, 2005b. The floating leg given in the equation above is the one correctly priced by the VIX^2 index (after rescaling), as shown by Schneider and Trojani, 2015a. Its continuous-time limit is:

$$\int_t^{t+\tau} \sigma_s^2 ds + 2 \sum_{t<s\leq t+\tau} \phi_0' \left(\frac{F_{t+s}}{F_{t+s-}} \right). \tag{3.3.3}$$

$p = 1$: Gamma swap

For $p = 1$ the CGF slope swap corresponds to the Gamma swap (Lee, 2010), also referred to as a Kullback-Leibler swao (Schneider and Trojani, 2015a). The price and payoff of are, respectively:

$$2\mathbb{E}_t^{\mathbb{Q}} \left[\frac{F_{t+\tau}}{F_t} \ln F_{t+\tau}/F_t \right]$$

$$2F_{t+\tau}/F_t \ln F_{t+\tau}/F_t - 2\sum_{j=1}^{N} \frac{F_{t_j} - F_{t_{j-1}}}{F_{t_{j-1}}} = \sum_{j=1}^{N} F_{t_j}/F_t \ln F_{t_j}/F_{t_{j-1}} - \frac{F_{t_j} - F_{t_{j-1}}}{F_{t_{j-1}}}.$$

The payoff's continuous-time limit is, after appropriate scaling:

$$2\int_t^{t+\tau} \frac{F_{t+s}}{F_t} \sigma_s^2 ds + 2 \sum_{t<s\leq t+\tau} \frac{F_{t+s-}}{F_t} \phi_1' \left(\frac{F_{t+s}}{F_{t+s-}} \right), \tag{3.3.4}$$

that is, it accumulates quadratic variation weighted by the cumulative return on the asset.

$p = 1/2$: Hellinger skew swap

For $p = 1/2$, the CGF slope swap is a rescaled version of the Hellinger skew swap found in Schneider and Trojani, 2015a, with price and payoff:

$$-\frac{8}{e^{K_t(p,\tau)}} \mathbb{E}_t^{\mathbb{Q}} \left[\left(\frac{F_{t+\tau}}{F_t} \right)^{\frac{1}{2}} \ln F_{t+\tau}/F_t \right],$$

$$-\frac{8}{e^{K_t(p,\tau)}} \sqrt{\frac{F_{t+\tau}}{F_t}} \ln F_{t+\tau}/F_t + \frac{8}{e^{K_t(p,\tau)}} \sum_{j=1}^{N} \frac{F_{t_j} - F_{t_{j-1}}}{F_{t_{j-1}}}$$

$$= -\frac{8}{e^{K_t(p,\tau)}} \sum_{j=1}^{N} \phi_{\frac{1}{2}}' \left(\frac{F_{t_{j-1}}}{F_t} \right) \left(\phi_{\frac{1}{2}} \left(\frac{F_{t_{j-1}}}{F_t} \right) - \frac{F_{t_j} - F_{t_{j-1}}}{2F_{t_{j-1}}} \right) + \phi_{\frac{1}{2}} \left(\frac{F_{t_{j-1}}}{F_t} \right) \left(\phi_{\frac{1}{2}}' \left(\frac{F_{t_{j-1}}}{F_{t_j}} \right) - \frac{F_{t_j} - F_{t_{j-1}}}{2F_{t_{j-1}}} \right.$$

The continuous-time limit of the payoff is:

$$\int_t^{t+\tau} \frac{\phi'_{\frac{1}{2}}\left(\frac{F_{t+s-}}{F_t}\right)}{e^{K_t(p,\tau)}}\sigma_s^2 ds - \frac{8}{e^{K_t(p,\tau)}} \sum_{t\leq s\leq t+\tau} \left[\phi_{\frac{1}{2}}\left(\frac{F_{t+s-}}{F_t}\right)\phi'_{\frac{1}{2}}\left(\frac{F_{t+s-}}{F_{t+s}}\right) + \phi'_{\frac{1}{2}}\left(\frac{F_{t+s-}}{F_t}\right)\left(\phi_{\frac{1}{2}}\left(\frac{F_{t+s-}}{F_{t+s}}\right) - 1\right)\right] \tag{3.3.5}$$

The payoff accumulates – in the first term – continuous quadratic variation with a positive or negative weight, depending on whether $F_{t_{j-1}}$ is smaller or greater than the starting value F_t. Additionally, in the second term, it accumulates second-order jump variation, with a similar weighting rule.

3.3.1 CGF slope swaps in AJDs

Equation (3.2.7) is a counterpart of (3.2.3) in affine jump diffusion settings. It follows immediately that the prices of CGF slope contracts are affine in the state variables in this class of models.

Proposition 7 (affine contracts)**.** *The price of a CGF slope swap at p is equal to:*

$$\frac{\mathbb{E}_t^{\mathbb{Q}}\left[\phi'_p\left(\frac{F_{t+\tau}}{F_t}\right)\right]}{e^{K_t(p,\tau)}} = \alpha'(p,\tau) + \beta'(p,\tau)\cdot v_t,$$

where the coefficients α', β' are obtained by finding the transform in Proposition 3 in Duffie, Pan, and Singleton, 2000.

This fact has been known and exploited for long for $p = 0$ (the variance swap). Note that in Section 3.3 the payoffs were rescaled for an easier interpretation: the CGF slope at $p = 0$ ($p = 1/2$) is, to first order, exposed to the negative of the second (third) moment of returns.

As summarized in Section 3.1, unspanned skewness (i.e. stochastic skewness that is not driven by factors impacting stochastic volatility) is a primary feature of option implied volatility surfaces which are heavily skewed at the short end. In many dimension-reduced estimation attempts in the literature, the term structure of variance swaps is used, alongside the stock return, as an observable statistic summarising option information. In what follows, I focus on the term structures of the variance, Hellinger skew and Gamma swaps in example models, and analyse the information about state variable values contained therein.

I build example models starting from the SVJCD2 specification in Bates, 2000. The base specification is denoted M1. The three remaining specifications modify the base model to introduce various degrees of unspanned skewness. Model parameters

are given in Table 3.1.

$$\begin{aligned}
\frac{dF_t}{F_{t-}} &= -\bar{\nu}_t dt + \phi_1\sqrt{v_{1t}}dW_{1t}^S + \phi_2\sqrt{v_{2t}}dW_{2t}^S + \int_{\mathbb{R}\setminus\{0\}}\left(e^{g(x,y)}-1\right)\nu_t(dx,dy)dt \\
dv_{1t} &= \kappa_1(\eta_1 - v_{1t})dt + \sigma_1\sqrt{v_{1t}}dW_{1t}^v + \int_{\mathbb{R}_+} y\nu_t(dx)dt \\
dv_{2t} &= \kappa_2(\eta_2 - v_{2t})dt + \sigma_2\sqrt{v_{2t}}dW_{2t}^v \\
dW_{it}^S dW_{it}^v &= \rho_i dt,
\end{aligned}$$

M1 through M3: $\frac{\nu_t(dx,dy)}{dx} = \frac{\lambda_1 v_{1t} + \lambda_2 v_{2t}}{\sqrt{2\pi}} e^{-\frac{(x-\mu_J)^2}{2\sigma_J^2}}$, while in M4: $\frac{\nu_t(dx,dy)}{dxdy} = \frac{\lambda_1 v_{1t}}{\rho_J \mu_J} e^{\frac{y}{\rho_J \mu_J}}$

name	M1	M2	M3	M4
η_1	0.0148	0.0148	0.0148	0.0148
κ_1	3.21	3.21	3.21	3.21
σ_1	0.24	0.24	0.24	0.24
ρ_1	-1	–	–	–
η_2	0.0195	0.0195	0.0195	0.0195
κ_2	0.85	0.85	0.85	0.85
σ_2	0.18	0.18	0.18	0.18
ρ_2	-0.314	0	-0.314	-0.314
λ_1	88.6	88.6	60.6	28.6
λ_2	–	–	15.2	–
μ_{JV}	–	–	–	0.05
μ_J	-0.054	-0.054	-0.054	–
ρ_{JV}	–	–	–	-0.99
σ_J	0.102	0.102	0.102	–
$\kappa_1^{\mathbb{Q}}$	2	2	2	2
$\kappa_2^{\mathbb{Q}}$	0.87	0.87	0.87	0.87
ϕ_1	1	0	0	0
ϕ_2	1	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{2}$

TABLE 3.1: Example model parameters

In the base specification (M1), both variance factors impact the level of the Brownian volatility of the stock, but only v_1 drives the jump intensity. Simultaneously, the innovations to both factors are negatively correlated with corresponding innovations to the forward stock price. Under the parameter choices, in all models the (primary) jump-driving factor v_1 mean-reverts quicker than v_2. In specifications M1 through M3, the jump component only impacts returns, and apart from a negative mean, the jumps have symmetric Gaussian tails. In specification M4, jumps in factor v_1 are exponential and feed into the stock price process. In the supplementary specification M2:

- I remove the intensity driver's impact on Brownian asset volatility, setting $\lambda_1 = 0$;
- remove v_2's leverage effect, setting $\rho_2 = 0$;

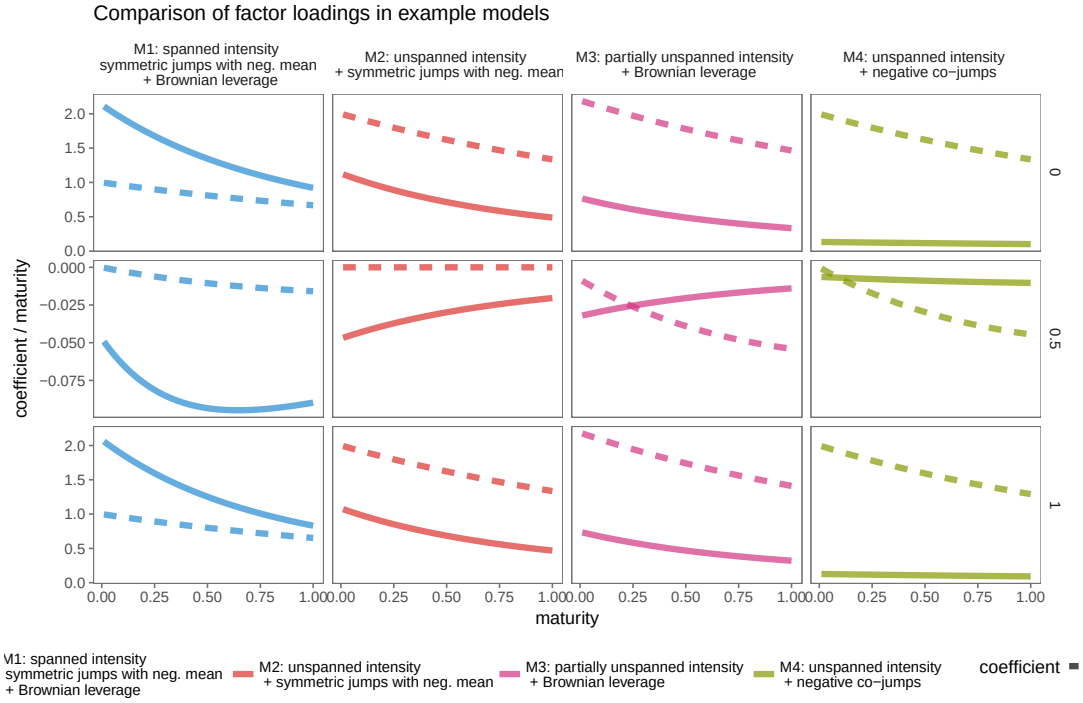


FIGURE 3.3: Scaled factor loadings in affine contracts, $\beta_j'(p,\tau|v_t)/\tau$ against time to maturity in two-factor models.

- I increase the second factor's impact on Brownian volatility, setting $\phi_2 = \sqrt{2}$

. Specification M3 retains the changes from M2 and:

- I set $\lambda_2 = 15.2$ so that short-term skewness is influenced by both factors, but Brownian volatility only by v_2.

Specification M4 retains the features of M2, but alters the jump distribution:

- the asset price and v_1 co-jump, and the jumps are exponentially distributed self-exciting, i.e. $\lambda_1 = 28.6$, $\mu_{JV} = 0.05$, and the jump leverage parameter is set to $\rho_J = -0.99$;

Additionally, in specifications M2 through M4 I increase the remaining factor's contribution to Brownian volatility so as to achieve a similar level of total return variance as in M1.

Figure 3.3 contains plots of (maturity-scaled) factor loadings in affine contracts $\beta_j'(p,\tau)/\tau$, as given in Proposition 7. Loadings on the VIX variance swap ($p = 0$), the Hellinger skewness swap ($p = 1/2$), and the Gamma swap ($p = 1$) (top-to-bottom) in the aforementioned models (left-to-right) are compared. Loadings on v_1 – unspanned by volatility dynamics in models M2 throgh M4 – are given in a solid line; on v_2 (volatility driver) in a dashed line.

While the plot is not a study of a general case, the changes in specification are rich enough to generate a diverse range of loading term structure behaviours. Note that the loadings of the variance and Gamma swap are almost identical in all models.

This shifts the focus to the differences in variance and Hellinger swaps. The relative loadings of the variance swap rates vary little across the term structure, especially in models with unspanned intensity of symmetric jumps (M1 and M2), where the loadings are approximately collinear; their shapes are similar, too, across models. To the contrary, the value of loadings for the Hellingers skew swap vary with maturity in significantly different ways than loadings of variance swaps. Across models, the skew swap loadings also vary significantly more than the variance swap loadings. The Hellinger swap loadings' common feature is that at short maturites, the loadings on factor v_2 converge to 0 whenever it does not drive jump intensity. If additionally this factor does not generate a leverage effect, the price of the Hellinger swap does not depend on it at all. Finally, model M4 shows an example where the factor v_1 that drives asymmetric, self-exciting co-jumps, is absent from the term structure of variance swap prices.

Ex-ante, the econometrician does not know the characterization of factor loadings in the data. With recent empirical evidence pointing to the importance of unspanned skewness, the takeaway of this section is that using short-term Hellinger skew swaps in estimation data offers new statistics about latent volatility and skewness drivers, with richer information content than the term structure of variance swaps. Furthermore, in certain extreme situations, the variance swap term structure can have a loading of essentially 0 on the isolated skewness driver, and Helliger skew swaps become essential for uncovering its $\mathbb{Q}$-dynamics.

3.3.2 Spanning in option markets

Carr and Madan, 2001 showed that in the presence of complete option markets, that is if European options on the underlying are available for strikes $K \in [0, \infty]$, an arbitrary payoff function $\pi(x)$, non-linear in the forward return, is available for trading in the following sense:

$$\pi(F_{t+\tau}) - \pi(F_t) - \pi'(F_t)(F_{t+\tau} - F_t) = \int_0^\infty \pi''(K_t) O(K_t, F_{t+\tau}) dK_t. \tag{3.3.6}$$

In the equation above, $O(K_t, F_{t+\tau})$ is the payoff, at time $t+\tau$, of a maturing option. The option is a call (put) if at time t, K_t was greater (smaller) than F_t (i.e. it's an out of the money option). The equation states that the delta-hedged increment [3] of function π can be obtained as a payoff of a portfolio of options at their maturity.

Incomplete markets

Equation (3.3.6) is an unobtainable idealization. Investors in real markets face two impediments: the option strike grid is discretized, and (more importantly), bounded. Their impact on the ability to hedge non-linear contracts is two-fold: discretization

[3] **Bregman1967** divergence, as noted by Schneider and Trojani, 2015a

causes inaccuracies in the non-linear hedge if the final asset value stays within the strike grid, while boundedness limits the ability to hedge extreme events. Boundedness also implies that the replicating portfolio's price is only an approximation of the true contract price, and the error uncertainty is unaccounted for. Such extreme events, however, happen extremely rarely, as discussed in the appendix to Martin, 2013 and by Orłowski, 2015. Martin also discusses the underpricing for variance contracts in the context of an implementation of 3.3.6 by simple discretization. Orłowski, 2015 shows that there are better alternatives, and optimal static replications can be chosen at any point in time, as to minimize expected hedging error. I use this method for calculating contract swap prices in the subsequent simulation exercises.

The model-based theoretical quantities are thus approximated as values of option portfolios):

$$K_t'(p,\tau) = \alpha'(p,\tau) + \beta'(p,\tau)\cdot v_t \approx \sum_{j=1}^{N} w_j^* O_\tau(K_{t,j}), \tag{3.3.7}$$

where w_j^* are portfolio weights that minimize expected hedging error. The choice of method for estimating $K_t'(p,\tau)$ from option data can have important implications for model estimation results. If a method such as (3.3.7) is chosen, the approximate contract prices are no longer linear in state variables. For certain ranges of factor values and availability of option contracts, the estimates can vary wildly from their theoretical CGF slope counterparts. While Martin, 2013 discusses possible biases when estimating the prices of variance swaps and concludes they would be small, they might be much more significant for Hellinger skew swap contracts, which load much stronger on far out of the money options than variance contracts. This is illustrated in Figure 3.4, which plots the relative approximation error of the (3.3.7) method. Other methods of estimating $K_t'(p,\tau)$ involve approximating the integral in (3.3.6) through estimating the implied volatility curve (and thus option prices), and extrapolating it both towards 0 and ∞. Such approaches introduce biases which are difficult to quantify and account for, as they are equivalent to making a statement about the specification of the jump distribution.

I account for potential biases in the filtering and estimation by constructing the filter directly on replicating portfolios,[4] rather than on model-implied theoretcial prices of variance and Hellinger skew swaps. I find that the departures from linearity are small enough that the Unscented Kalman Filter (Wan and Merwe, 2000; Merwe and Wan, 2001) is a formidable tool for the task at hand (for applications in Finance, see for example Li, 2013 and Christoffersen et al., 2014).

[4] I find the prices of the replicating portfolios in the model in the follwing sense: if in the option data, J strikes $\{K_{tj}^*\}_1^J$ are available, and the optimal replicating portfolio for a given swap rate at time t assigns weights $\{w_{tj}^*\}_1^J$ to the options, then in the filter I construct the model-implied price of the portfolio by finding the prices of options with the given strikes, and forming the portfolio with the optimal weights estimated in the data.

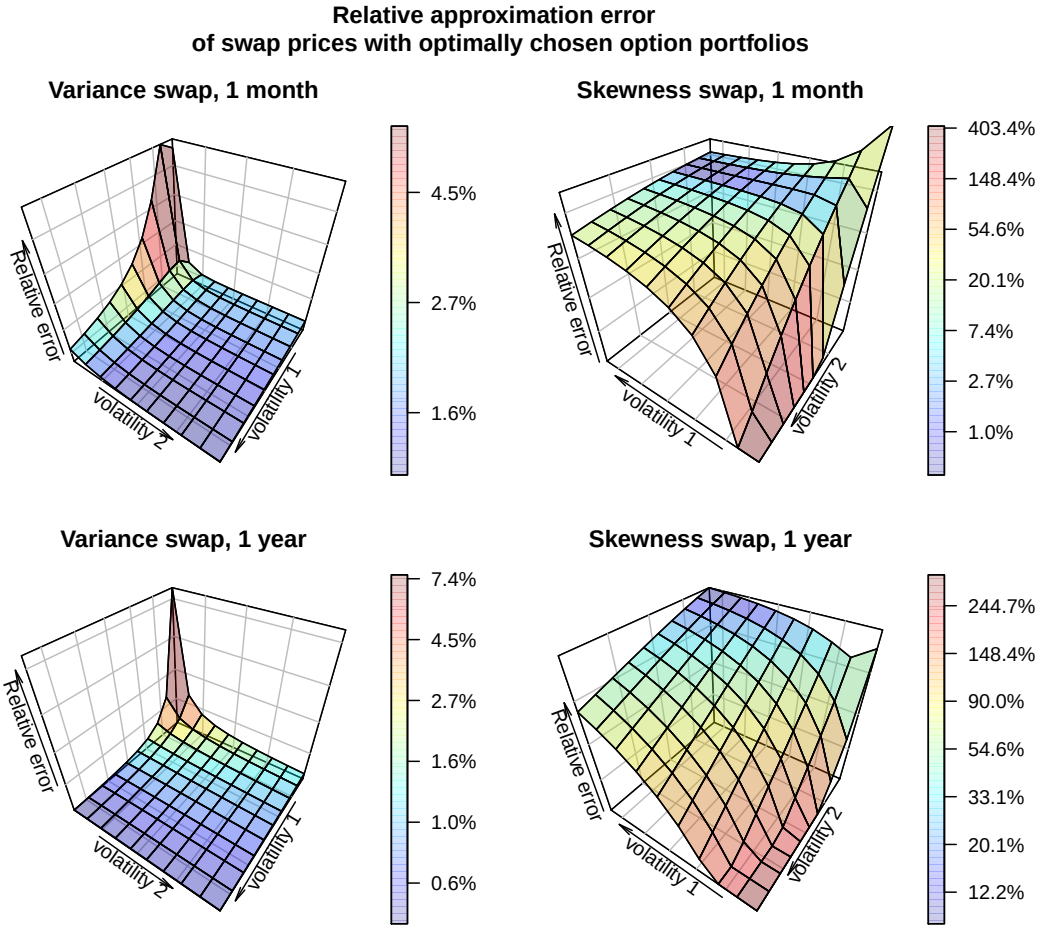


FIGURE 3.4: Model M4: relative approximation error $100\%times\left(\frac{p_{\text{true}}-p_{\text{portfolio}}}{p_{\text{true}}}\right)$ of variance and skewness swap prices with option portfolios chosen with the method described in Orłowski, 2015. Curvature in plotted surfaces represents departures from linearity of approximate contract price with respect to the factors. Extreme overpricing of skewness swaps is possible at very low values of factor v_1.

3.4 Simulation study

In the simulation study I focus on the consequences of estimation dataset choice for state variable recovery. I look at three aspects of the results. First, I examine the filtering error directly. The error feeds into further quantities of interest, such as factor premia and option prices. Thus, I translate the filtering error into option pricing error. This setup abstracts from parameter estimation error and allows to quantify signal fidelity losses that are solely due to factors under the econometrician's control, i.e. the composition of the dataset, while keeping the filtering method constant.

I scrutinize models M1 and M4, with specification as in Section 3.3.1 and parameters as in Table 3.1. The three datasets for filtering are:

1. **opts**: for each date, the dataset contains the prices of 30 options per maturity, at 1- and 12-month maturities;
2. **vs**: for each date, the dataset contains the prices of the 1- and 12-month variance swaps;
3. **hs+vs**: for each date, the dataset contains the prices of the 1-month Hellinger skew swap and 12-month variance swap.

The prices of Hellinger skew and variance swaps in the two reduced-dimension datasets are calculated with the options available in dataset **opts** and in line with remarks in Section 3.3.2. In each dataset I assume that the signal-to-noise ratio is known, and I set the standard deviation of noise to 10% of a contract price's standard deviation. The noise is independent across contracts (be it swaps or options). Note that this assumption puts the reduced-dimension filters at a disadvantage, because independent noise in individual options would likely largely cancel out in portfolio formation.

3.4.1 State filtering

Two of the dataset setups – **vs** and **hs+vs** – reduce the option surface observed at each time point to a set of summary statistics. **Opts**, on the other hand, uses the complete price set for the same maturities. All datasets are noisy. I demonstrate, with the use of simulations, that the **hs+vs** dataset is no worse a base for the purpose of state recovery than the complete **opts** dataset.

Filtered factor values for models M1 and M4 are plotted in Figures 3.5 and 3.7, respectively. In M1, both factors are recovered with similar accuracy when using the **hs+vs** and **opts** datasets. With **vs**, recovery significantly deteriorates. A summary of filtering error distribution for model M1 is given in Figure 3.6. For factor v_1, which drives both volatility and jump intensity, recovery with **hs+vs** and **opts** is approximately on par, whereas with **vs** the interquantile range of recovery error increases two-fold. For factor v_2, the **opts** dataset outperforms both **hs+vs** and **vs**, albeit **hs+vs** still reduces the interquantile range of recovery error by approximately 50% with respect to focusing on variance swaps. Model M4 is of particular interest

– in this setup, v_1 drives strictly negative jumps the asset price, which correspond to jumps in volatility, as inspired by the recent empirical literature (Andersen, Todorov, and Fusari, 2015). For v_2 the results are similar to those in model M1. Filtering v_1, on the other hand, poses an insurmountable challenge to a filter based on **vs**. It also suffers with the **opts** dataset. The path of v_1 filtered from **vs** seems to bear little relation to the true path, with a correlation coefficient of approximately 30%. State recovery with this dataset is equally bad across all factor levels. The **opts** based filter correctly recovers the v_1 path when v_1 does not stray far from its long-run mean, however the filtered values are extremely noisy after v_1 jumps, when its value is very high. To the contrary, the filter based on **hs+vs** does not suffer from such extreme errors. Figure 3.8 plots filtering errors in model M4, stratified by the respective factor quantiles. When factor v_1 is below its 75th percentile, filtering with both the **opts** and **hs+vs** dataset yields reasonable results. When v_1 values rise, however, filtering with the whole option surface yields very noisy estimates – the interquantile range is equal to that for the variance swap dataset, and more extreme errors are recorded; the estimates seem nonetheless unbiased, contrary to those recovered with **vs**.

Across the whole paths of v_1 and v_2, using the **hs+vs** dataset reduces mean absolute filtering error by 57% and 44%, respectively, in model M1, and by 80% and 23% in model M4. In the absence of unspanned skewness (M1) the **hs+vs** dataset offers filtering performance equivalent to that of the full option panel in **opts**: it improves the recovery of v_1 by 20%, but hinders the recovery of v_2 by 23%. In the setting with significant unspanned skewness the **hs+vs** dataset, which compactifies the information in the volatility surface into two informative signals, significantly improves state recovery not only with respect to the usual **vs** approach, but also with respect to the whole implied volatility surface, the unfavourable observation noise assumption notwithstanding. Absolute filtering errors of v_1 and v_2 are reduced by 55% and 13%, respectively.

3.4.2 Option pricing

Section 3.4.1 gives a good indication of filtering performance gains when Hellinger skew swaps are introduced to the dataset. Filtering quality translates directly to a model's utility for inference about quantities such as option prices, and the importance of any state recovery improvements should be judged through the lens of the task at hand. From the point of view of an econometrician who estimates an asset pricing model, there are three sources of error in estimates of option prices: the filtering error, estimation error (understood as the distance between the estimated and (pseudo-)true parameter values), and model specification error. I concentrate on the former: in models with known parameters, similarly to the previous section, I investigate the impact of the filtering dataset specification on estimated option prices. I

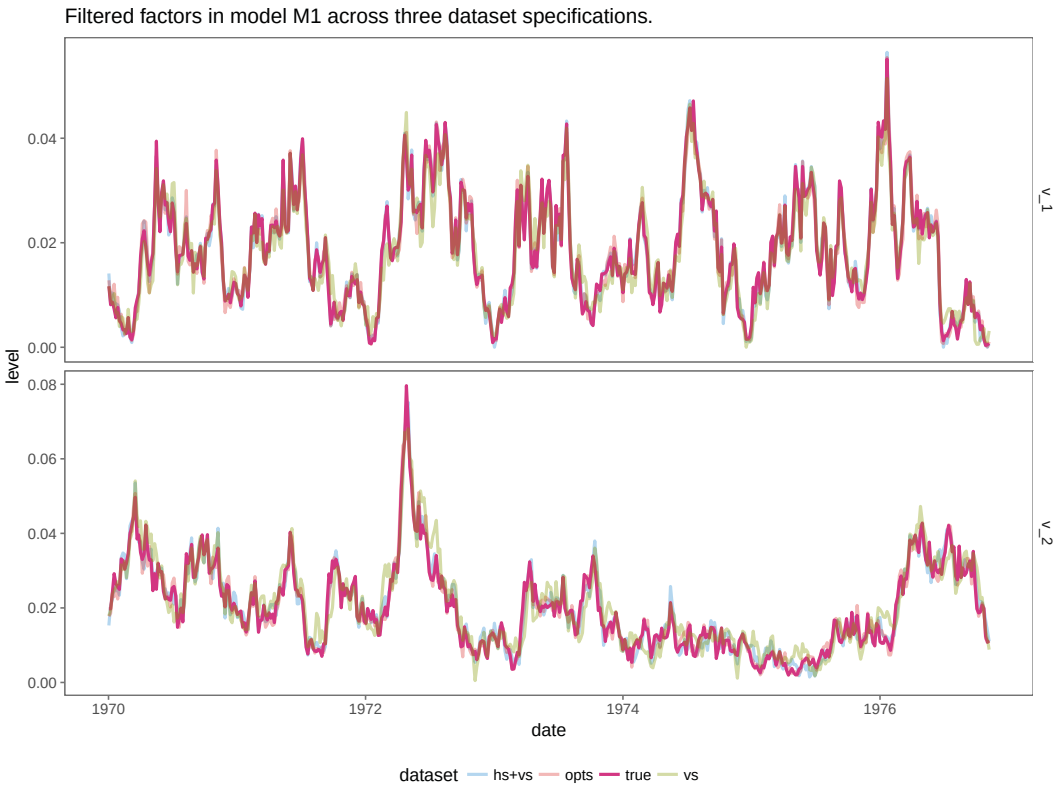


FIGURE 3.5: Filtered factor paths in model M1 with three dataset setups: full option surface, **opts**; variance swaps, **vs**; Hellinger skew and variance swaps, **hs+vs**. For each factor, errors are grouped by the quartile of the factor's level.

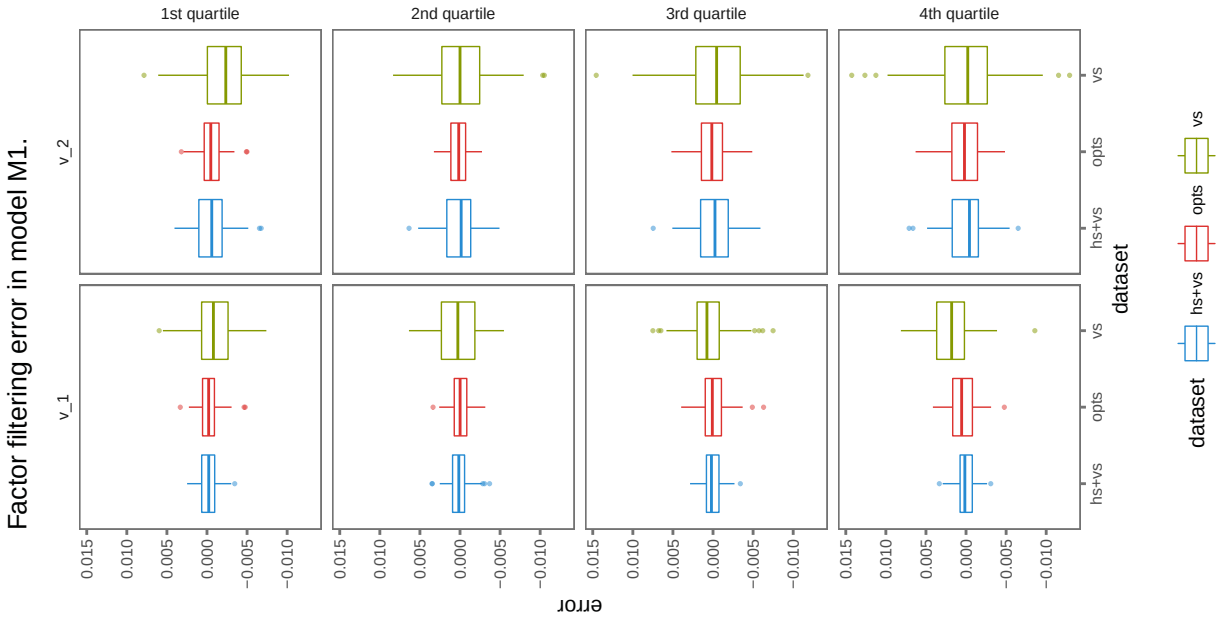


FIGURE 3.6: Factor filtering error in model M1 with three three dataset setups: full option surface, **opts**; variance swaps, **vs**; Hellinger skew and variance swaps, **hs+vs**.

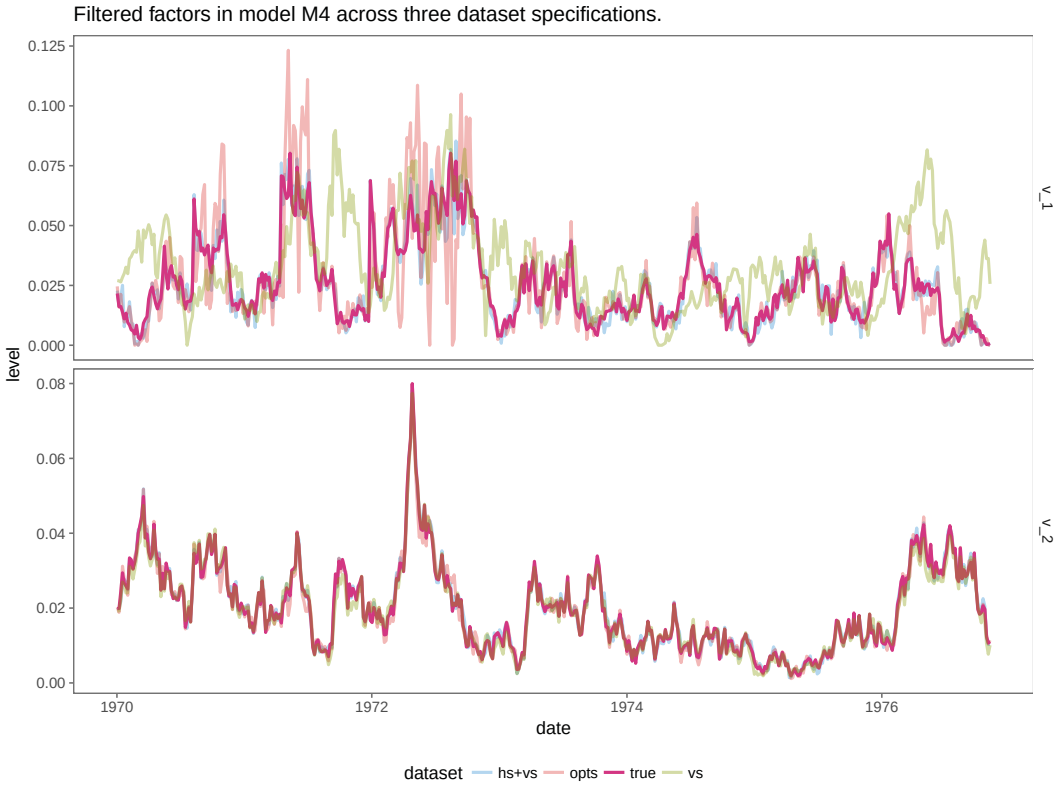


FIGURE 3.7: Filtered factor paths in model M4 with three dataset setups: full option surface, **opts**; variance swaps, **vs**; Hellinger skew and variance swaps, **hs+vs**.

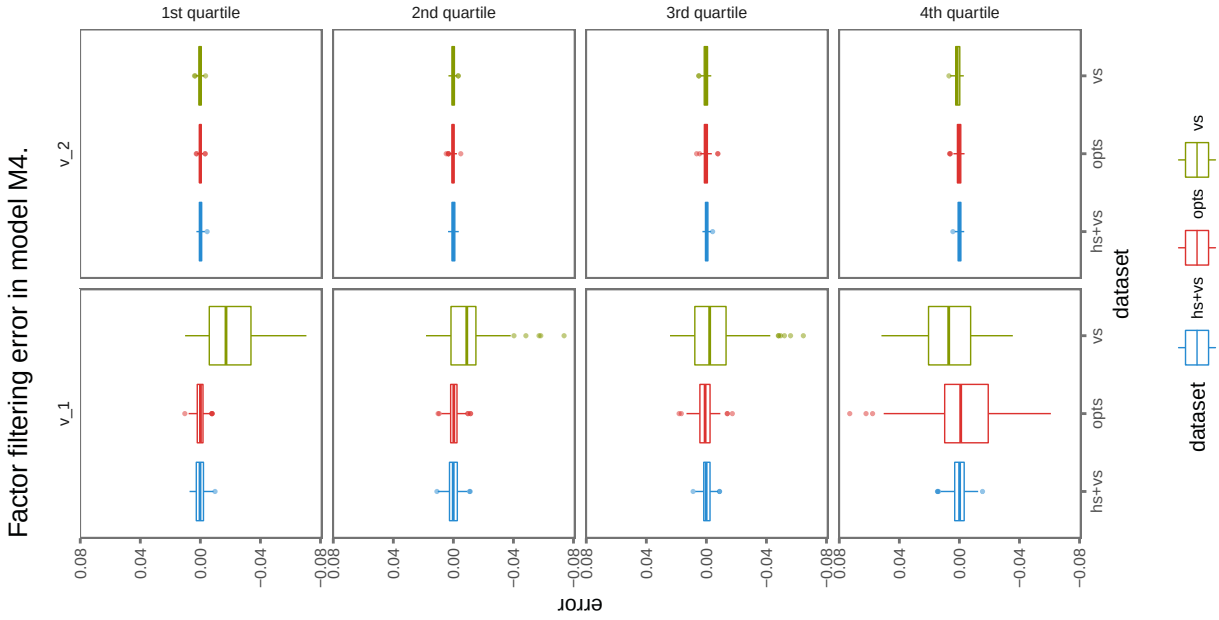


FIGURE 3.8: Factor filtering error in model M4 with three three dataset setups: full option surface, **opts**; variance swaps, **vs**; Hellinger skew and variance swaps, **hs+vs**. For each factor, errors are grouped by the quartile of the factor's level.

judge the setups in terms of relative error of an option's implied volatility,

$$\epsilon_t(k,\tau) := \frac{IV_{\text{true}}(k,\tau) - IV_{\text{filtered}}(k,\tau)}{IV_{\text{true}}(k,\tau)}. \tag{3.4.1}$$

The option panels used for evaluation span maturities from 1 week to 1 year, with 30 options per maturity. The state variables are filtered from dataset which contain options with only two maturities: 1 month and 1 year. I include the single-week maturing options, because of the rising importance of the short-term options market, and because they are the most informative of jump risk determinants. Andersen, Fusari, and Todorov, 2016 showed that the dynamics of "weeklies'" prices is indeed highly informative of tail risks prevailing in the index option market. I visualize the distributions of relative pricing error, and the values of mean absolute pricing error in Figures 3.9 through 3.12.

Distributions of option pricing errors are displayed in Figures 3.9 (model M1) and 3.10 (model M4). The options are split into categories across three dimensions: option maturity, option moneyness level (defined as $\frac{\ln K/F}{\sqrt{\tau} IV_{ATM}}$), and the quartiles of prevailing implied volatility skew for options of given maturity. Options are grouped in moneyness buckets in the following manner: the label denotes that the bucket contains options with moneyness levels smaller than the label, i.e. for 0 the bucket contains puts with moneyness levels $(-2, 0]$. The simulated option surfaces contain more puts than calls, as is typical in observed panels. Implied volatility skew is calculated as the difference between median implied volatility of options in the $(-4, -2]$ moneyness bucket and at the money implied volatility, for every time stamp and maturity. In model M1, implied skew is strongly positively correlated with factor v_1 across all maturities and moneyness groups (correlations ranging from 0.8 to 0.95), and not correlated with factor v_2 (correlations ranging from -0.2 to 0.2). In model M4 the implied volatility skew is less correlated with factor v_1, even if this factor is the primary driver of time variation in the third moment of returns (correlations range between 0.5 and 0.8). Factor v_2, on the other hand, is strongly negative correlated with the IV skew (-0.8 to -0.5).

There are important differences in the range of pricing errors resulting from modifying the filtering dataset. Certain regularities are clearly visible for both model parametrizations: pricing errors decrease with option maturity (due to factor mean-reversion: the present factor value has little impact on long-maturity options), and for implied volatility skew levels above the median and maturities above 1 month, call options and at the money put options offer a greater challenge than out of the money put options. Most importantly, for most maturities and implied volatility skew levels, the **vs** dataset yields the noisiest option price estimates. Judged in terms on mean absolute percentage pricing error (MAPE), the **vs** dataset is the worst perfomer across the board, as seen in Figures 3.11 and 3.12.

The performance differences of filters based on **hs+vs** and **opts** datasets are more nuanced. In terms of MAPE (Figures 3.11 and 3.12), in model M4 the **hs+vs** dataset

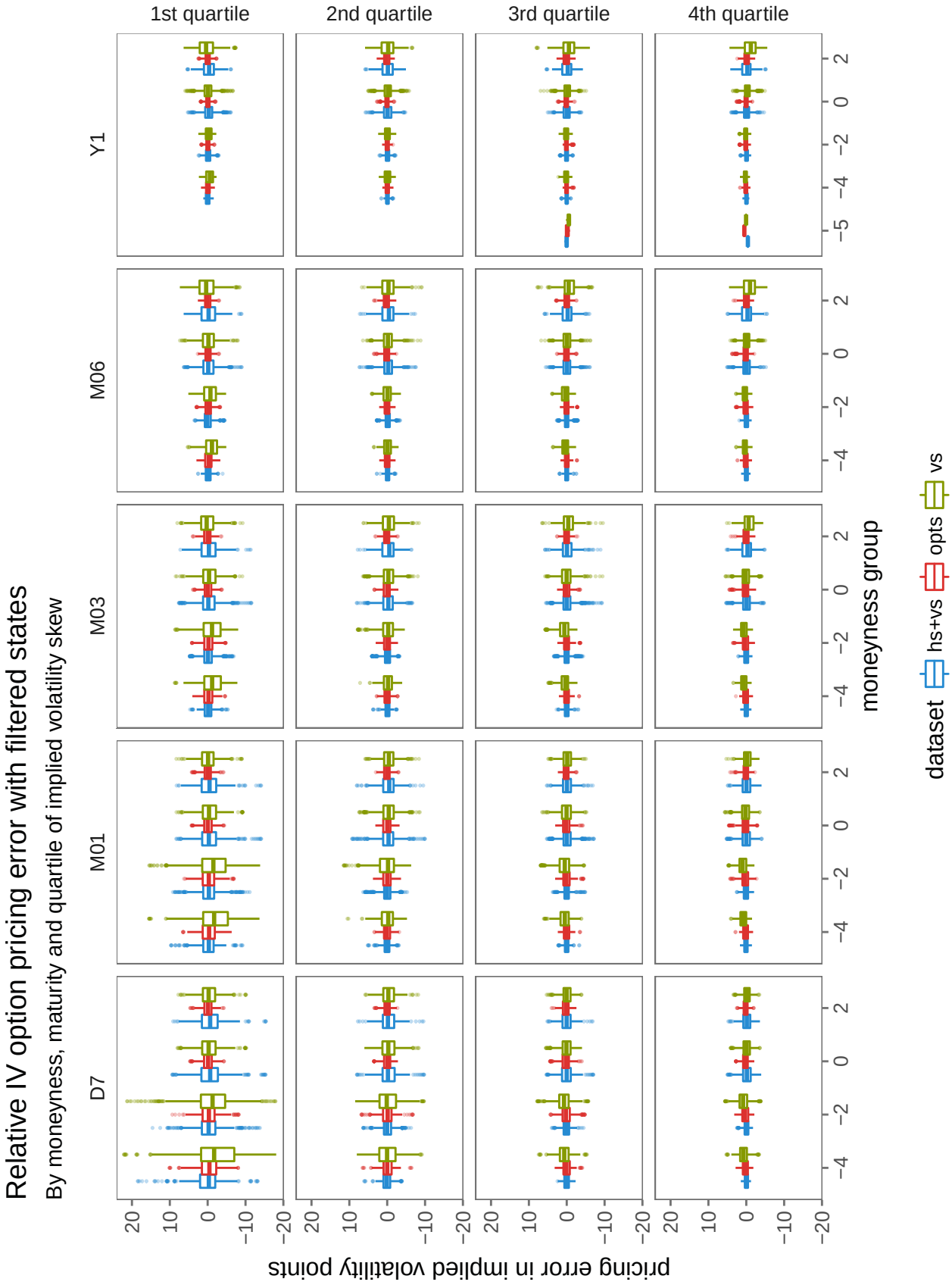


FIGURE 3.9: Option pricing error in model M1 when parameters are known, but state variables are filtered. Error is expressed in percentage of a given option's true implied volatility, $\frac{IV_{\text{true}}(k,\tau)-IV_{\text{filtered}}(k,\tau)}{IV_{\text{true}}(k,\tau)}$. Pricing errors are grouped by option moneyness levels defined as $k = \frac{\ln K/F}{\sqrt{\tau} IV_{\text{ATM}}}$, and by the level of implied volatility skewness.

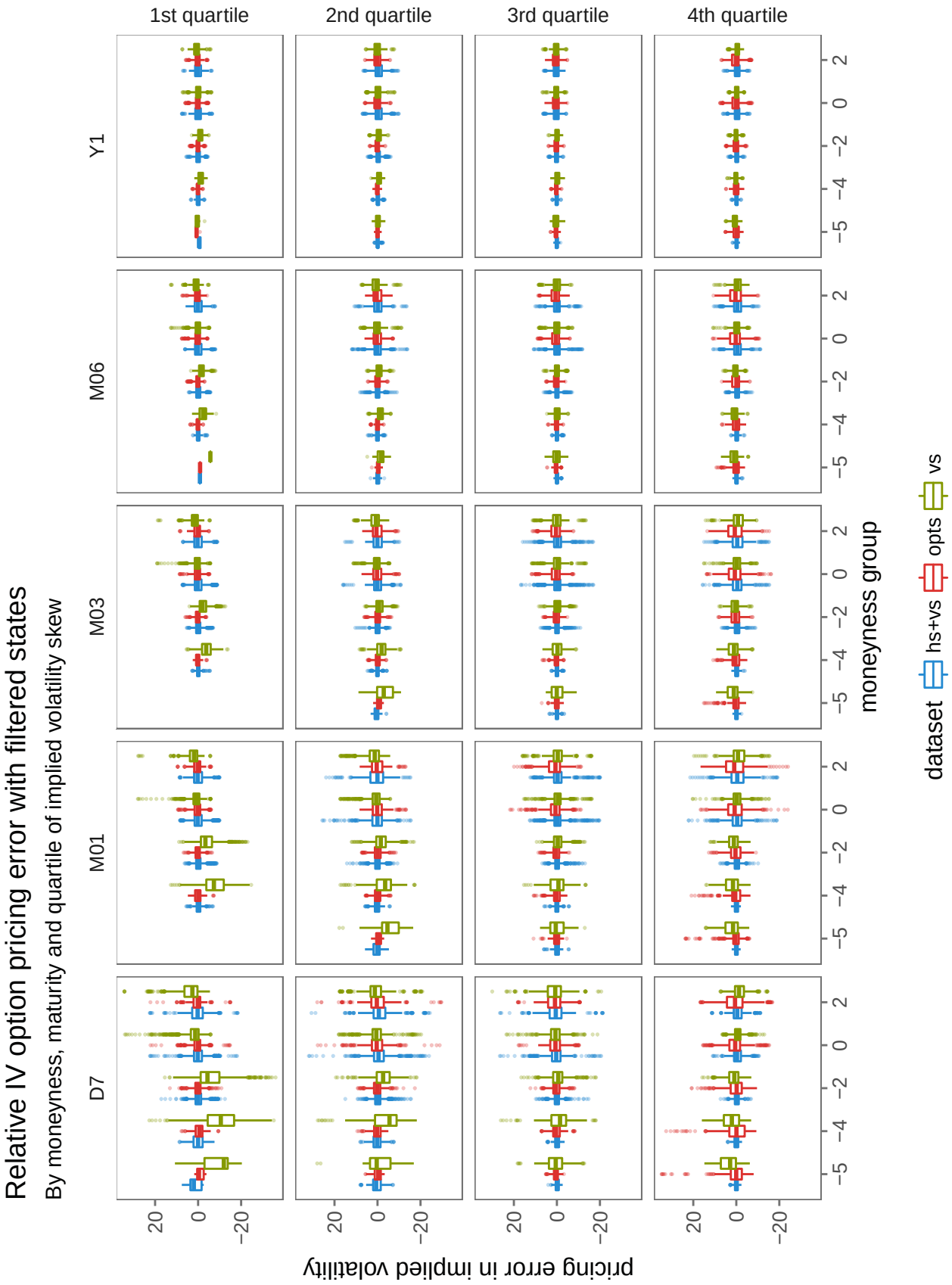


FIGURE 3.10: Option pricing error in model M4 when parameters are known, but state variables are filtered. Error is expressed in percentage of a given option's true implied volatility, $\frac{IV_{\text{true}}(k,\tau)-IV_{\text{filtered}}(k,\tau)}{IV_{\text{true}}(k,\tau)}$. Pricing errors are grouped by option moneyness levels defined as $k = \frac{\ln K/F}{\sqrt{\tau} IV_{\text{ATM}}}$, and by the level of implied volatility skewness.

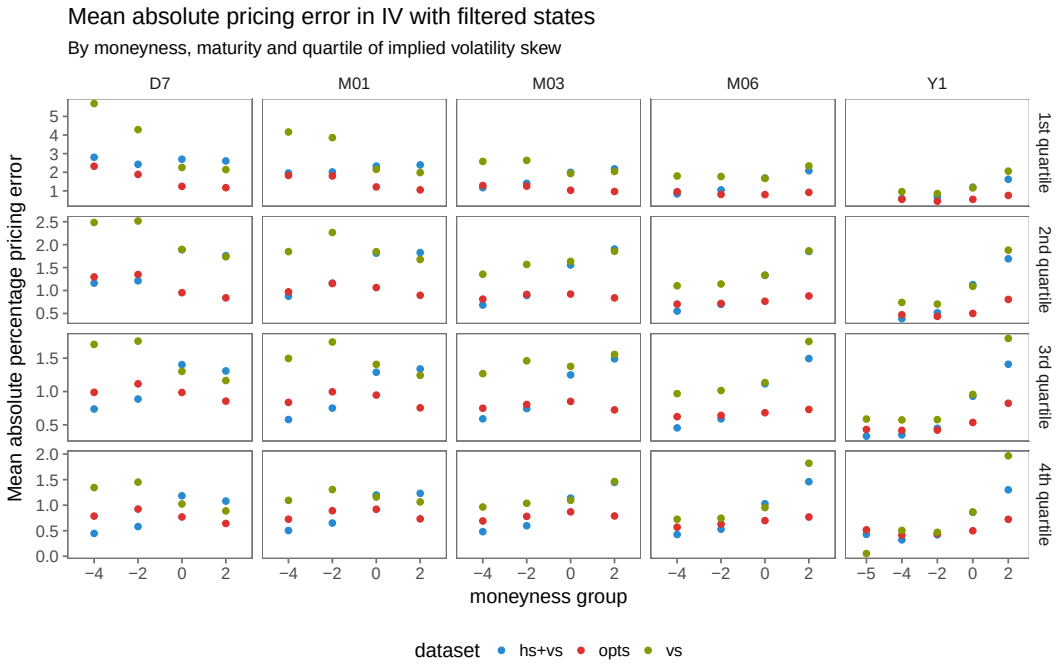


FIGURE 3.11: Mean absolute percentage pricing error in model M1 when parameters are known, but state variables are filtered. Error is expressed in percentage of a given option's true implied volatility, $\frac{IV_{\text{true}}(k,\tau)-IV_{\text{filtered}}(k,\tau)}{IV_{\text{true}}(k,\tau)}$. Pricing errors are grouped by option moneyness levels defined as $k = \frac{\ln K/F}{\sqrt{\tau}IV_{\text{ATM}}}$, and by the level of implied volatility skewness.

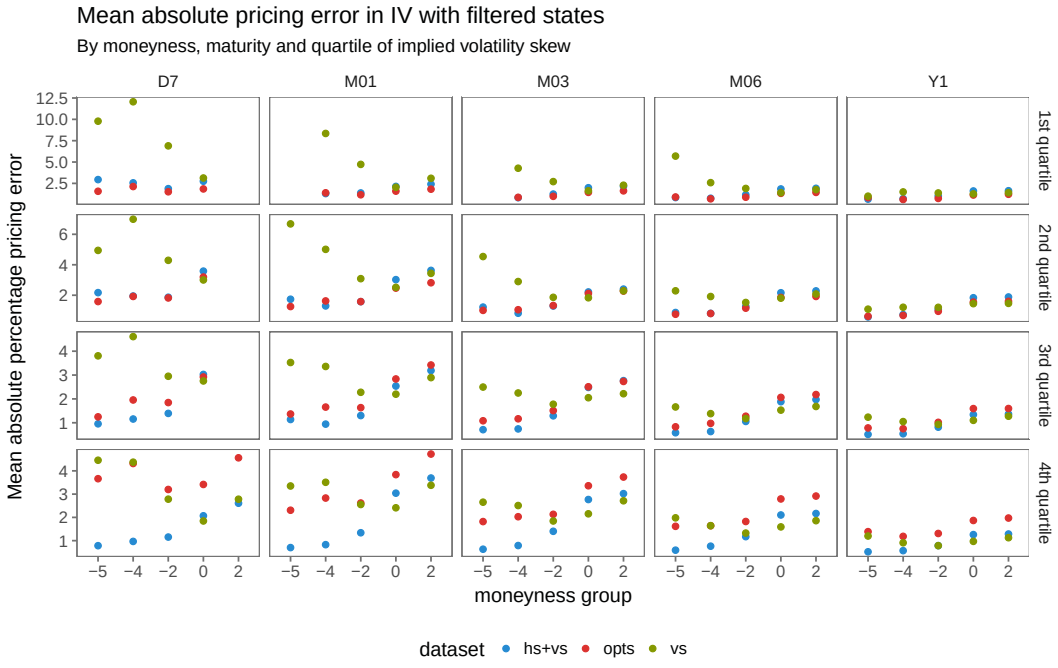


FIGURE 3.12: Mean absolute percentage pricing error in model M4 when parameters are known, but state variables are filtered. Error is expressed in percentage of a given option's true implied volatility, $\frac{IV_{\text{true}}(k,\tau)-IV_{\text{filtered}}(k,\tau)}{IV_{\text{true}}(k,\tau)}$. Pricing errors are grouped by option moneyness levels defined as $k = \frac{\ln K/F}{\sqrt{\tau}IV_{\text{ATM}}}$, and by the level of implied volatility skewness.

offers an improvement over **opts**, with MAPEs of 1.76% and 1.94%, respectively. In model M1, the **opts** dataset retains an advantage, with a MAPE of 0.87% vs **hs+vs**'s 1.22%. Performance gains in model M4 are most significant when the implied volatility skew is above its 75th percentile, for put options with maturities up to 3 months. Interquantile ranges of pricing errors are reduced by 50%, and outlying pricing errors are practically eliminated. The filter based on the **opts** dataset performs better when pricing call options, which are practically not impacted by the dynamics of factor v_1 in model M4.

As I argued in the introduction, recent literature implies that a model specification in the spirit of M4 is more relevant for empirical research than such as M1. A filter design which replaces short-maturity variance swaps with Hellinger skew swaps offers an improved identification of the path of a skewness-driving factor in a model such as M4, not only with respect to a variance-swap based filter, but also with respect to employing the whole option surface. The improvements in filtering translate into smaller option pricing errors when the model parameters are known. While a general recommendation for filter design is beyond the scope of this study, I present preliminary evidence that a dimension-reduced filter can perform on par or better than a more computationally involved alternative. The discussion in this paper also abstracts from the specifications of observation noise. It is self-evident that proposing an observation noise structure for an option panel with hundreds of observed assets opens more room for error than specifying one for a small number of swap rate contracts.

3.5 Conclusions

The econometrician who faces the task of estimating an option pricing model when state variables are unobservable, must specify the equations of the no-arbitrage pricing model, and the blueprint for the filter to uncover the latent variables. Assuming a given filtering method (a particle filter, or an UKF, as in this paper), the econometrician's task is limited to choosing the right observed quantities, which will yield good state recovery properties. This is much akin to the problem of sensor placement in the control literature, e.g. Tzoumas, Jadbabaie, and Pappas, 2016: using signals from more sensors is not always optimal; a sensor's signal to noise ratio changes over time, and when it is low, the sensor should be excluded from the observation set. Historically, the empirical option pricing literature resolved to use either all sensors – complete option panels, with multiple maturities and hundreds of options per each data point – or certain linear combinations of sensor signal – prices of variance swaps, which can be interpreted as portfolios of options.

I demonstrated that the variance swap observation design performs significantly worse in terms of both state filtering and, as a consequence, option pricing with filtered states, than the complete option panel dataset. In a next step, I showed how replacing short-term variance swaps with Hellinger skew contracts in filter design

makes the reduced-dimension filter competitive with one based on the whole option surface. In a simulated setting with rich volatility and jump intensity dynamics, which mimics features of factor dynamics uncovered by the recent literature, this augmented filtering system outperforms a filter based on the complete option surface. Furthermore, I characterized the variance swap and Hellinger skew contracts in the context of a broader class of CGF slope swaps. The swap rates in this class of contracts are linear in the unobserved latent states when affine jump diffusion dynamics is assumed.

The question of optimal filter design in option pricing model estimation remains open. I leave to extensions of this paper an empirical application of hereby developed methods and the problem of state-dependent observation equation design.

A Appendices

A.1 Appendix to Chapter 1

A.1.1 Divergence: useful expressions

Higher-order measures

The following equations express the derivatives of $F_s^{-p} G_p(F_t, F_s)$ with respect to p expressed in terms of $y := \ln F_t/F_s$.

$$\frac{\partial}{\partial p}\frac{D_p(F_t,F_s)}{F_s^p} = \frac{ye^{py}-(e^y-1)}{p(p-1)} - \left(\frac{1}{p^2(p-1)}+\frac{1}{p(p-1)^2}\right)\left(e^{py}-1-p\left(e^y-1\right)\right), \tag{A.1.1}$$

$$\lim_{p\to 0}\frac{\partial}{\partial p}\frac{D_p(F_t,F_s)}{F_s^p} = -\frac{y^2}{2} - y + \left(e^y-1\right), \tag{A.1.2}$$

$$\lim_{p\to 1}\frac{\partial}{\partial p}\frac{D_p(F_t,F_s)}{F_s^p} = \frac{y^2e^y}{2} - ye^y + \left(e^y-1\right), \tag{A.1.3}$$

$$\frac{\partial^2}{\partial p^2}\frac{D_p(F_t,F_s)}{F_s^p} = \frac{y^2e^{py}}{p(p-1)} - 2\left(\frac{1}{p^2(p-1)}+\frac{1}{p(p-1)^2}\right)\left(ye^{py}-(e^y-1)\right) \tag{A.1.4}$$

$$+\left(\frac{2}{p^3(p-1)}+\frac{2}{p^2(p-1)^2}+\frac{2}{p(p-1)^3}\right)\left(e^{py}-1-p(e^y-1)\right),$$

$$\lim_{p\to 0}\frac{\partial^2}{\partial p^2}\frac{D_p(F_t,F_s)}{F_s^p} = -\frac{y^3}{3} - y^2 - 2y + 2(e^y-1), \tag{A.1.5}$$

$$\lim_{p\to 0}\frac{\partial^2}{\partial p^2}\frac{D_p(F_t,F_s)}{F_s^p} = \frac{y^3e^3}{3} - y^2e^y + 2ye^y - 2(e^y-1). \tag{A.1.6}$$

Higher-order tradability

The following equations express the quantities required for jump replication as portfolios of weighted divergences. I start with defining additional functions to facilitate notation.

$$\psi_{p,k}(x) := \frac{x^p \ln^k x}{p(p-1)} \qquad \Psi_{p,k}(y,x) := \psi_{p,k}(y) - \psi_{p,k}(x) - \psi'_{p,k}(x)(y-x)$$

The function Ψ is simply the Bregman divergence of ψ. I also define $\psi_{0,k}(x) := \ln^k x$ and $\psi_{1,k}(x) := x\ln^k x$. The following equations express the scaled divergence

derivatives as divergences to which equation (1.3.11) is applicable.

$$\frac{\partial}{\partial p}\frac{D_p(F_t,F_s)}{F_s^p} = \frac{\Psi_{p,1}(F_t,F_s)}{F_s^p} - \frac{\ln F_s + \left(\frac{1}{p}+\frac{1}{p-1}\right)}{F_s^p} D_p(F_t,F_s) \tag{A.1.7}$$

$$\lim_{p\to 0}\frac{\partial}{\partial p}\frac{D_p(F_t,F_s)}{F_s^p} = -\frac{\Psi_{0,2}(F_t,F_s)}{2} - (1-\ln F_s)\, D_0(F_t,F_s) \tag{A.1.8}$$

$$\lim_{p\to 1}\frac{\partial}{\partial p}\frac{D_p(F_t,F_s)}{F_s^p} = \frac{\Psi_{1,2}(F_t,F_s)}{2F_s} - \frac{\ln F_s - 1}{F_s} D_1(F_t,F_s) \tag{A.1.9}$$

$$\begin{aligned}\frac{\partial^2}{\partial p^2}\frac{D_p(F_t,F_s)}{F_s^p} &= \frac{\Psi_{p,2}(F_t,F-s)}{F_s^p} - 2\frac{\ln F_s + \left(\frac{1}{p}+\frac{1}{p-1}\right)}{F_s^p}\Psi_{p,1}(F_t,F_s) \\ &+ \frac{\ln^2 F_s + 2\left(\frac{1}{p}+\frac{1}{p-1}\right)\left(\ln F_s + \frac{1}{p}+\frac{1}{p-1}\right) - \frac{2}{p(p-1)}}{F_s^p} D_p(F_t,F_s)\end{aligned} \tag{A.1.10}$$

$$\begin{aligned}\lim_{p\to 0}\frac{\partial^2}{\partial p^2}\frac{D_p(F_t,F_s)}{F_s^p} &= -\frac{\Psi_{0,3}(F_t,F_s)}{3} + (\ln F_s - 1)\,\Psi_{0,2}(F_t,F_s) \\ &+ \left(\ln^2 F_s - 2(\ln F_s - 1)\right) D_0(F_t,F_s)\end{aligned} \tag{A.1.11}$$

$$\begin{aligned}\lim_{p\to 1}\frac{\partial^2}{\partial p^2}\frac{D_p(F_t,F_s)}{F_s^p} &= \frac{\Psi_{0,3}(F_t,F_s)}{3F_s} - \frac{\ln F_s + 1}{F_s}\Psi_{0,2}(F_t,F_s) \\ &+ \frac{\ln^2 F_s + 2(\ln F_s + 1)}{F_s} D_1(F_t,F_s)\end{aligned} \tag{A.1.12}$$

A.1.2 Option portfolio weights

In this section I give a detailed description of the quadratic programming problem of optimal static replication. Rewrite the optimization criterion in equation (1.4.3) more compactly and separate the terms, with $g(x)$ the function to be replicated, $\eta(x)$ a weighting function and $O(x,K_j)$ the payoff of an out-of-the money option:

$$\begin{aligned}&\int_0^\infty \eta(x)\left(g(x) - \sum_{j=1}^J w_j O(x,K_j)\right)^2 dx \\ &= \int_0^\infty \eta(x)\left(g(x)^2 + \left(\sum_{j=1}^J w_j O(x,K_j)\right)^2 - 2g(x)\sum_{j=1}^J w_j O(x,K_j)\right) dx \\ &= C + \int_0^\infty \eta(x)\left(\sum_{j=1}^J\sum_{k=1}^J w_j w_k O(x,K_j)O(x,K_k) - 2g(x)\sum_{j=1}^J w_j O(x,K_j)\right) dx \\ &= C + w^T\mathbf{Q}w - \mathbf{q}^T w\end{aligned}$$

Above, $\mathbf{Q}$ is a square $J \times J$ matrix with entry Q_{ij}:

$$Q_{ii} = \int_0^\infty \eta(x) O(x, K_i)^2 dx = \begin{cases} \int_0^{K_i} \eta(x)(K_i - x)^2 dx & \text{if } K_i \leq F \\ \int_{K_i}^\infty \eta(x)(x - K_i)^2 dx & \text{if } K_i > F \end{cases}$$

$$Q_{ij} = \int_0^\infty \eta(x) O(x, K_i) O(x, K_j) dx$$

$$= \begin{cases} 0 & \text{if } K_i < F < K_j \text{ or } K_j < F < K_i \\ \int_0^{\min(K_i, K_j)} \eta(x)(x - K_i)(x - K_j) dx & \text{if } K_i, K_j \leq F \\ \int_{\max(K_i, K_j)}^\infty \eta(x)(K_i - x)(K_j - x) dx & \text{if } K_i, K_j > F \end{cases},$$

$\mathbf{q} \in \mathbb{R}^J$ with entries q_j:

$$q_j = \int_0^\infty 2\eta(x) g(x) O(x, K_j) dx = \begin{cases} \int_0^{K_j} 2\eta(x) g(x)(x - K_j) dx & \text{if } K_j \leq F \\ \int_{K_j}^\infty 2\eta(x) g(x)(K_j - x) dx & \text{if } K_j > F \end{cases},$$

while $C = \int_0^\infty \eta(x) g(x)^2 dx$ is a constant that can be ignored in the optimization. All integrals can be easily evaluated numerically, and in special cases, for example with $\eta(x) \equiv \mathbf{1}_{\{K_1 - \varepsilon \leq x \leq K_J + \varepsilon\}}$, analytically.

The matrix $\mathbf{Q}$ is block-diagonal and block-wise positive definite. Thus for computational efficiency and without loss of replication accuracy, it's possible to split the above problem into two convex subproblems for replicating the payoff function $g(x)$ separately for $x \leq F$ and $x > F$.

A.1.3 Data filters

I remove the following records from my dataset of all SPX option quotes and trades:

1. Options which are not those with monthly settlements effective on the morning of the third Friday of the month (e.g. weeklies and LEAPS), and which are not the closest maturity;
2. Quotes with zero bid prices;
3. Quotes with non-positive bid-ask spreads;
4. Quotes with bid-ask spread greater than 500% of the bid price;
5. Options with strikes whose standardised moneyness, $\log(K/F)/(\sigma_{IV}^{ATM}\sqrt{T-t})$ is lower than -12 or higher than 6;
6. Quotes whose mid-prices are not contained in the bid-ask spread after imposing the no-arbitrage mid-price system of the following subsubsection A.1.3.

No-arbitrage option price system

The relative price of a put option is the integral of the risk-neutral distribution function of the return on the underlying asset:

$$p_{K,T} \equiv \frac{P_{K,T}}{F} = \int_0^K P(F_T/F \leq x)dx, \tag{A.1.13}$$

and as such, for $K_j > K_i$, I should have $p_{K_j,T} - p_{K_i,t} > 0$. Based on this relation I construct an L^1 correction method for observed relative option prices which ensures that the resulting mid-prices do not allow for arbitrage, and that does not rely on parametric assumptions. Let J denote the number of option price quotes prevailing at a given point in time. Let $\hat{\pi}_{K,T}$ denote the price which ensures there is no arbitrage in the system. I find $\hat{\pi} = [\hat{\pi}_{K_1,T}, \ldots, \hat{\pi}_{K_J,T}]^T$ as:

$$\hat{\pi} = \underset{\pi \in \mathbb{R}_+^N}{\operatorname{argmin}} \sum_{j=1}^{J} \left(p_{K_j,T} - X_{(j,\cdot)}\pi\right) \left(\frac{1}{2} - \mathbf{1}_{\{p_{K_j,T} - X_{(j,\cdot)}\pi \leq 0\}}\right)$$

with

$X_{(\cdot,1)} = \mathbf{1}_{J\times 1}$

$X_{(i,j)} = K_i - K_{j-1}$ for $i \in 2, \ldots, J$ and $j \in 2, \ldots, i$.

Applying the above to mid-prices $p_{K,T}$ and subsequently discarding these records for which $\hat{\pi}_{K_j,T}$ is outside the bid-ask spread allows us to find a set of no-arbitrage "mid" prices which would not allow for arbitrage in a market free of transaction profits.

A.1.4 High frequency option trading in the Black-Scholes model

The simplest Brownian-driven model of stock price behavimy allows us to learn about risk compensation in benchmark cases. First, the model is put-call symmetric with constant volatility. Second, asset price paths are continuous. Hence, at the high-frequency trading limit the realized divergence strategies will pay exactly $\sigma^2(t_n - t_0)$ while the skewness and quarticity strategies have zero payoffs. Furthermore, in the known case of the log contract, i.e. $D_n^0(F)$, where divergence weights are constant, the risk premium for divergence is 0.

Assume the following dynamics of the forward price of a stock index under $\mathbb{P}$:

$$dF_s = \mu F_s ds + \sigma F_s dW_s\,, \tag{A.1.14}$$

and under $\mathbb{Q}$ (where F is a martingale):

$$dF_s = \sigma F_s dW_s\,. \tag{A.1.15}$$

The statistical and risk-neutral characteristic functions of the log-return on F are, respectively:

$$\mathbb{E}_s^{\mathbb{P}}\left[e^{u\log F_t/F_s}\right] = e^{(t-s)\left(\frac{1}{2}\sigma^2(u^2-u)+\mu u\right)} \text{ and } \mathbb{E}_s^{\mathbb{Q}}\left[e^{u\log F_t/F_s}\right] = e^{\frac{1}{2}\sigma^2(t-s)(u^2-u)}. \tag{A.1.16}$$

I assume that the trading strategy is executed on a period $[0,T]$, where T is the maturity of the options and forward contracts.

Dynamics of some functions of the underlying in the Black-Scholes model

Here I gather expressions that are useful in the following sections when studying the trading profits in the Black-Scholes model:

$$\phi_s(p,T-s) \equiv \mathbb{E}_s^{\mathbb{Q}}\left[e^{u\log F_t/F_s}\right] - 1, \tag{A.1.17}$$

$$\frac{dF_s^p}{F_s^p} = \left(p\mu + \frac{p(p-1)}{2}\sigma^2\right)ds + p\sigma dW_s, \tag{A.1.18}$$

$$\frac{dF_s^p \log F_s}{F_s^p} = \left(p\mu\log F_s + \mu + \frac{1}{2}\sigma^2\left[p(p-1)\log F_s + (2p-1)\right]\right)ds + \sigma\left(p\log F_s + 1\right)dW_s \tag{A.1.19}$$

Realized divergence trading

I proceed with deriving the expressions for the payoff, trading profits, expected tradings profits, and finally the premium for trading realized divergence in the Black-Scholes model. First recall that equation (1.3.8) implies that the payoff $D_n^p(F)$ of the options and forwards at maturity reaches the limit of integrated divergence:

$$D_n^p(F) \xrightarrow{\mathbb{P}} \frac{1}{2}\sigma^2 T \equiv D_\infty^p(F). \tag{A.1.20}$$

From equation (1.3.13) I can decompose the costs of obtaining the payoff, $\mathcal{C}\left[D_\infty^p\right]$, as:

$$\mathcal{C}\left[D_\infty^p(F)\right] = \frac{\phi_0(p,T-s)}{p(p-1)} - \int_0^T \frac{\phi_s(p,T-s)}{p(p-1)}\frac{dF_s^p}{F_s^p}.$$

The conditional premium for trading is $DP_{\mathbb{Q}_T}(p) \equiv \mathbb{E}_0^{\mathbb{P}}\left[D_\infty^p(F)\right] - \mathbb{E}_0^{\mathbb{Q}_T}\left[\mathcal{C}\left[D_\infty^p(F)\right]\right]$ if the market allows for contingent contracts such that the profits of rebalancing can be contracted at time 0. The conditional premium in case where the profits of rebalancing have to be borne along the trading path, is: $DP_{\mathbb{P}}(p) \equiv \mathbb{E}_0^{\mathbb{P}}\left[D_\infty^p(F)\right] - \mathbb{E}_0^{\mathbb{P}}\left[\mathcal{C}\left[D_\infty^p(F)\right]\right]$. $DP_{\mathbb{Q}_T}(p)$ is not identically 0:

$$DP_{\mathbb{Q}_T}(p) = \sigma^2 T - 2\frac{\phi_0(p,T)}{p(p-1)}, \quad \lim_{p\to 0} DP_{\mathbb{Q}_T}(p) = \lim_{p\to 1} DP_{\mathbb{Q}_T}(p) = 0. \tag{A.1.21}$$

$DP_{\mathbb{Q}_T}(p) > 0$ for $p \in (0,1)$, $DP_{\mathbb{Q}_T}(p) < 0$ for $p \notin [0,1]$. The convexity adjustment for $p \notin \{0,1\}$ arises because the second derivative of the price of divergence at $p = 0$ and

$p = 1$ is polynomial in σ^2, but has an exponential term for $p \notin \{0, 1\}$. The conditional premium in case where rebalancing cannot be contracted at time 0, depends on the risk premium μ:

$$DP_{\mathbb{P}}(p) = \sigma^2 T + \frac{\mu T}{p-1} - \frac{2\mu\phi_0(p,T)}{p(p-1)^2\sigma^2} - \frac{2\phi_0(p,T)}{p(p-1)}, \tag{A.1.22}$$

$$\lim_{p\to 0} DP_{\mathbb{P}}(p) = 0 \text{ and } \lim_{p\to 1} DP_{\mathbb{P}}(p) = -\frac{\mu\sigma^2 T^2}{4}. \tag{A.1.23}$$

From (A.1.18) and (A.1.20) I infer that $\mathcal{C}\left[D^p_\infty(F)\right]$ is normally distributed, with mean and variance given below:

$$\mathcal{C}\left[D^p_\infty(F)\right] \sim \mathcal{N}\left(-\frac{(2\mu + (p-1)\sigma^2)\left(\frac{2\phi_0(p,T)}{(p-1)p\sigma^2} - T\right)}{2(p-1)}, \left(-2\frac{\phi_0(p,T)}{p(p-1)^2\sigma} + \frac{T\sigma}{p-1}\right)^2\right). \tag{A.1.24}$$

On $p \in [0, 1]$ the convexity adjustment profit is an order of magnitude smaller than the expected rebalancing profit.

Realized skewness trading

Trading jump skewness in the Black-Scholes model reveals more peculiar features of dynamic option trading. Recall that from equation (1.3.9), the payoff $S^p_n(F)$ of the options and forwards at maturity reaches the limit:

$$S^p_n(F) \xrightarrow{\mathbb{P}} 0 \equiv S^p_\infty(F), \tag{A.1.25}$$

because the path of F is continuous. Analogously to equation (1.3.13), I can decompose the cost of obtaining the payoff, $\mathcal{C}\left[S^p_\infty\right]$, as:

$$\begin{aligned}
\mathcal{C}\left[S^p_\infty\right] &= \frac{\phi_0'(p,T)}{p(p-1)} - \frac{2p-1}{(p(p-1))^2}\phi_0(p,T) - \int_0^T \frac{\phi_s(p,T-s)}{p(p-1)}\frac{dF^p_s \log F_s}{F^p_s} - \int_0^T \frac{\phi_s'(p,T-s)}{p(p-1)}\frac{dF^p_s}{F^p_s} \\
&+ \int_0^T \frac{\log F_s \phi_s(p,T-s)}{p(p-1)}\frac{dF^p_s}{F^p_s} + \int_0^T \frac{(2p-1)\phi_s(p,T-s)}{(p(p-1))^2}\frac{dF^p_s}{F^p_s} \\
&= \frac{\phi_0'(p,T)}{p(p-1)} - \frac{2p-1}{(p(p-1))^2}\phi_0(p,T) \\
&- \int_0^T \frac{\phi_s'(p,T)}{p(p-1)}\left(p\mu + \frac{p(p-1)}{2}\sigma^2\right)ds - \int_0^T \frac{\phi_s'(p,T)}{p(p-1)}p\sigma dW_s, \\
&- \int_0^T \frac{\phi_s(p,T-s)}{p(p-1)}\left(\mu + \frac{2p-1}{2}\sigma^2\right)ds - \int_0^T \frac{\phi_s(p,T-s)}{p(p-1)}\sigma dW_s \\
&+ \int_0^T \frac{(2p-1)\phi_s(p,T-s)}{(p(p-1))^2}\left(p\mu + \frac{p(p-1)}{2}\sigma^2\right)ds + \int_0^T \frac{(2p-1)\phi_s(p,T-s)}{(p(p-1))^2}p\sigma dW_s
\end{aligned}$$

The conditional premium for trading is $SP_{\mathbb{Q}_T}(p) \equiv \mathbb{E}^{\mathbb{P}}_0\left[S^p_\infty(F)\right] - \mathbb{E}^{\mathbb{Q}_T}_0\left[\mathcal{C}\left[S^p_\infty(F)\right]\right]$ if the market allows for contingent contracts such that the profits of rebalancing can be contracted at time 0. The conditional premium in case where the profits of

rebalancing have to be borne along the trading path, is: $SP_{\mathbb{P}}(p) \equiv \mathbb{E}_0^{\mathbb{P}}\left[S_\infty^p(F)\right] - \mathbb{E}_0^{\mathbb{P}}\left[\mathcal{C}\left[S_\infty^p(F)\right]\right]$. Note that the swap rate for skewness is zero if and only if $p = 1/2$. Similarly to the previous section, I can derive closed-form expressions for $SP_{\mathbb{Q}_T}(p)$ and $SP_{\mathbb{P}}(p)$, but in order to save space, I only present the limiting cases:

$$SP_{\mathbb{Q}_T}(p): \quad \lim_{p\to 0} SP_{\mathbb{Q}_T}(p) = \frac{\sigma^4 T^2}{4} \text{ and } \lim_{p\to 1} SP_{\mathbb{Q}_T}(p) = -\frac{\sigma^4 T^2}{4} \tag{A.1.26}$$

$$SP_{\mathbb{P}}(p): \quad \lim_{p\to 0} SP_{\mathbb{P}}(p) = \frac{\sigma^2\left(\sigma^2-\mu\right)T^2}{4} \text{ and } \lim_{p\to 1} SP_{\mathbb{P}}(p) = -\frac{\sigma^2\left(\sigma^2+\mu\right)T^2}{4} - \frac{\mu\sigma^4 T^3}{24}. \tag{A.1.27}$$

A.1.5 Tables

TABLE A.1: Trading and realised variation measures: replication accuracy. Weekly settlements.

RM	Min	$q_{0.25}$	Med	Mean	$q_{0.75}$	Max	SD	MAD	MAE	ME	p-val
Panel A: divergence											
$D_{1/2}$	0.00106	0.00423	0.00714	0.0119	0.0132	0.178	0.0158	0.00591			
$D_{1/2}$	-0.00215	0.004	0.00706	0.0117	0.0132	0.178	0.0157	0.00599	0.000266	0.000175	0.999
Panel B: skewness											
$S_{\gamma^J}(1/2)$	-472	-3.02	0.825	-1.14	5.21	392	57.8	5.99			
$S_{\gamma^J}(1/2)$	-472	-3.18	1.18	-1.19	7.97	363	61.3	7.85	5.52	0.0478	0.808
Panel C: quarticity											
$Q_{\gamma^J}(1/2)$	0.0872	1	3.84	32	15.2	1.73e+03	127	5.07			
$Q_{\gamma^J}(1/2)$	-1.09e+03	0.871	3.82	29.4	15.6	1.73e+03	177	5.2	21.2	2.59	0.808

Replication accuracy summary for divergence (Panel A), skewness (Panel B) and quarticity (Panel C) strategies associated with weighting γ^J. The first and third rows of each panel report the true realised measures calculated from forward prices. q_x denotes $100x$th percentile; Med denotes median; SD denotes standard deviation; MAD denotes median absolute deviation; . MAE denotes mean absolute error; ME denotes mean error.

TABLE A.2: Trading and realised variation measures: replication accuracy. Weekly settlements, 5-min hedging.

RM	Min	$q_{0.25}$	Med	Mean	$q_{0.75}$	Max	SD	MAD	MAE	ME	p-val
Panel A: divergence											
$D_{1/2}$	0.00106	0.00423	0.00714	0.0119	0.0132	0.178	0.0158	0.00591			
$D_{1/2}$	-0.00215	0.004	0.00706	0.0117	0.0132	0.178	0.0157	0.00599	0.000266	0.000175	0.999
Panel B: skewness											
$S_{\gamma^J}(1/2)$	-236	-1.51	0.413	-0.572	2.61	196	28.9	2.99			
$S_{\gamma^J}(1/2)$	-236	-1.59	0.591	-0.596	3.98	182	30.7	3.93	2.76	0.0239	0.808
Panel C: quarticity											
$Q_{\gamma^J}(1/2)$	0.0436	0.5	1.92	16	7.58	865	63.7	2.54			
$Q_{\gamma^J}(1/2)$	-546	0.435	1.91	14.7	7.82	865	88.6	2.6	10.6	1.29	0.808

Replication accuracy summary for divergence (Panel A), skewness (Panel B) and quarticity (Panel C) strategies associated with weighting γ^J. The first and third rows of each panel report the true realised measures calculated from forward prices. q_x denotes $100x$th percentile; Med denotes median; SD denotes standard deviation; MAD denotes median absolute deviation; . MAE denotes mean absolute error; ME denotes mean error.

TABLE A.3: Summary statistics of divergence trading with weeklies at one hour frequency

qty	Min	$q_{0.25}$	Med	CI (median) 90%		$q_{0.75}$	Max	Mean	SD	CI (mean) 90%	
Panel A: divergence											
$D_{\gamma^J}(1/2)$	0.001029	0.003974	0.006537	0.00562	0.007515	0.01272	0.1779	0.01125	0.01625	0.009885	0.01336
$\mathcal{C}[D_{\gamma^J}(1/2)]$	0.002626	0.006924	0.00997	0.009466	0.01068	0.01614	0.1258	0.01512	0.01586	0.01371	0.01705
$\mathcal{P}[D_{\gamma^J}(1/2)]$	-0.09832	-0.006246	-0.003358	-0.003731	-0.002765	-0.0006309	0.09479	-0.003864	0.01264	-0.00512	-0.002534
$\mathcal{T}[D_{\gamma^J}(1/2)]$	-0.1099	-0.0005735	0.002374	0.001949	0.00276	0.005057	0.09161	0.002348	0.01248	0.0008954	0.00347
Panel B: skewness											
$S_{\gamma^J}(1/2)$	-0.6292	-0.003806	0.0014	0.0006428	0.001934	0.006875	0.3067	-0.003219	0.06929	-0.01207	0.002467
$\mathcal{C}[S_{\gamma^J}(1/2)]$	-7.418	-0.2774	-0.1417	-0.1596	-0.1258	-0.0662	7.162	-0.239	0.8755	-0.3281	-0.1468
$\mathcal{P}[S_{\gamma^J}(1/2)]$	-7.791	0.06375	0.1479	0.1227	0.1641	0.2733	7.471	0.2357	0.9036	0.1401	0.3223
$\mathcal{T}[S_{\gamma^J}(1/2)]$	-10.75	-0.01135	0.08051	0.06109	0.09702	0.1746	5.907	0.0724	0.9869	-0.06306	0.1535
Panel C: quarticity											
$Q_{\gamma^J}(1/2)$	5.727e-06	0.0001766	0.0006414	0.0005001	0.0007549	0.002488	0.32	0.005494	0.02463	0.003719	0.00985
$\mathcal{C}[Q_{\gamma^J}(1/2)]$	-14.25	0.08055	0.1394	0.1273	0.1526	0.2984	14.43	0.303	1.675	0.1186	0.463
$\mathcal{P}[Q_{\gamma^J}(1/2)]$	-14.42	-0.2975	-0.1373	-0.1537	-0.1274	-0.08031	14.38	-0.2975	1.686	-0.4615	-0.1036
$\mathcal{T}[Q_{\gamma^J}(1/2)]$	-24.25	0.03526	0.07528	0.06325	0.08464	0.1364	10.31	0.05263	2.037	-0.2669	0.2065

Descriptive statistics of hourly frequency divergence trading profits with weekly option settlement for strategy γ^J. Trading period: from 2010-12-31 to 2015-12-24 with a total of 261 weeks. $\mathcal{C}[\cdot]$ denotes the integrated option cost at mid prices; $\mathcal{P}[\cdot]$ denotes total trading profits at mid prices; $\mathcal{T}[\cdot]$ denotes total trading profits after the transaction costs (bid-ask spread) are taken into account. $\mathcal{P}[\cdot]$ and $\mathcal{T}[\cdot]$ are reported in the following way: if mean $\mathcal{P}[\cdot]$ from a *long* strategy is negative, $\mathcal{P}[\cdot]$ and $\mathcal{T}[\cdot]$ from a *short* strategy are reported. q_x denotes $100x$th percentile; Med denotes median; SD denotes standard deviation; MAD denotes median absolute deviation.

TABLE A.4: Summary statistics of divergence trading with weeklies at five minute frequency

qty	Min	$q_{0.25}$	Med	CI (median) 90%		$q_{0.75}$	Max	Mean	SD	CI (mean) 90%	
Panel A: divergence											
$D_{\gamma^J}(1/2)$	0.001459	0.007246	0.0121	0.01065	0.01294	0.01905	0.2688	0.01818	0.02479	0.01613	0.02139
$\mathcal{C}[D_{\gamma^J}(1/2)]$	0.002819	0.00717	0.01031	0.009538	0.01064	0.016	0.1387	0.01541	0.01641	0.01395	0.01735
$\mathcal{P}[D_{\gamma^J}(1/2)]$	-0.05686	-0.003522	0.0009186	0.0003282	0.00158	0.005803	0.228	0.002774	0.01852	0.001339	0.005425
$\mathcal{T}[D_{\gamma^J}(1/2)]$	-0.06671	-0.00482	0.0004762	-0.0006722	0.001155	0.004921	0.2198	0.001267	0.01832	-0.0001882	0.003765
D_{VIX}	0.001459	0.007246	0.01209	0.01072	0.01294	0.01904	0.2703	0.01819	0.02485	0.01611	0.0213
$\mathcal{C}[D_{VIX}]$	0.00282	0.007213	0.01024	0.009548	0.01078	0.01607	0.1425	0.01551	0.0166	0.01406	0.01747
$\mathcal{P}[D_{VIX}]$	-0.05852	-0.003677	0.00087	0.0002947	0.001558	0.005733	0.2393	0.00268	0.01919	0.001179	0.00546
$\mathcal{T}[D_{VIX}]$	-0.06775	-0.004908	0.0004272	-0.0007096	0.001099	0.00492	0.2378	0.001265	0.01932	-0.0002345	0.004018
Panel B: skewness											
$S_{\gamma^J}(1/2)$	-0.2547	-0.004106	0.000539	0.0001603	0.0009623	0.004029	0.2361	-0.002871	0.04359	-0.007489	0.001221
$\mathcal{C}[S_{\gamma^J}(1/2)]$	-8.172	-0.2847	-0.1569	-0.1759	-0.1458	-0.07935	7.269	-0.2746	0.9273	-0.3714	-0.1834
$\mathcal{P}[S_{\gamma^J}(1/2)]$	-7.033	0.07723	0.158	0.1445	0.1717	0.2874	8.221	0.2717	0.9307	0.1773	0.3688
$\mathcal{T}[S_{\gamma^J}(1/2)]$	-13.82	-0.06591	0.04021	0.02469	0.05654	0.1201	4.915	-0.05125	1.179	-0.2232	0.0356
Panel C: quarticity											
$Q_{\gamma^J}(1/2)$	4.672e-06	0.0001729	0.0005	0.0003709	0.000607	0.00159	0.1813	0.003479	0.01398	0.002399	0.005714
$\mathcal{C}[Q_{\gamma^J}(1/2)]$	-22.91	0.08741	0.1436	0.1234	0.1537	0.2823	15.87	0.3187	1.986	0.06931	0.4879
$\mathcal{P}[Q_{\gamma^J}(1/2)]$	-15.86	-0.2819	-0.1431	-0.154	-0.1281	-0.08728	23.09	-0.3153	1.994	-0.4868	-0.06533
$\mathcal{T}[Q_{\gamma^J}(1/2)]$	-48.68	-0.003656	0.03855	0.0327	0.04237	0.09346	8.409	-0.154	3.283	-0.8162	0.06882

Descriptive statistics of 5-minute frequency divergence trading profits with weekly option settlement for strategy γ^J. Trading period: from 2010-12-31 to 2015-12-24 with a total of 261 weeks. $\mathcal{C}[\cdot]$ denotes the integrated option cost at mid prices; $\mathcal{P}[\cdot]$ denotes total trading profits at mid prices; $\mathcal{T}[\cdot]$ denotes total trading profits after the transaction costs (bid-ask spread) are taken into account. $\mathcal{P}[\cdot]$ and $\mathcal{T}[\cdot]$ are reported in the following way: if mean $\mathcal{P}[\cdot]$ from a *long* strategy is negative, $\mathcal{P}[\cdot]$ and $\mathcal{T}[\cdot]$ from a *short* strategy are reported. q_x denotes $100x$th percentile; Med denotes median; SD denotes standard deviation; MAD denotes median absolute deviation.

TABLE A.5: Summary statistics of open-to-close divergence trading with weeklies at five minute frequency

qty	Min	$q_{0.25}$	Med	CI (median) 90%		$q_{0.75}$	Max	Mean	SD	CI (mean) 90%	
Panel A: divergence											
$D_{\gamma^J}(1/2)$	0.0001151	0.002495	0.005041	0.004664	0.005358	0.01163	0.7204	0.0117	0.02824	0.01064	0.01344
$\mathcal{C}[D_{\gamma^J}(1/2)]$	-0.4276	0.001219	0.004199	0.003946	0.004414	0.008529	0.4149	0.006361	0.02443	0.005168	0.007492
$\mathcal{P}[D_{\gamma^J}(1/2)]$	-0.2294	-0.003282	0.0007545	0.0004215	0.001214	0.008023	0.6754	0.005334	0.03417	0.003959	0.00722
$\mathcal{T}[D_{\gamma^J}(1/2)]$	-0.3177	-0.009363	-0.003626	-0.004279	-0.003216	0.00263	0.5354	-0.003355	0.03102	-0.004775	-0.001795
D_{VIX}	0.0001155	0.002493	0.005041	0.004672	0.005365	0.01163	0.7201	0.0117	0.02823	0.01066	0.01342
$\mathcal{C}[D_{VIX}]$	-0.4516	0.001125	0.00416	0.003936	0.00437	0.008577	0.4274	0.006352	0.02529	0.005148	0.007567
$\mathcal{P}[D_{VIX}]$	-0.2418	-0.003288	0.0007952	0.0003824	0.001168	0.007879	0.6746	0.005344	0.03478	0.003961	0.007314
$\mathcal{T}[D_{VIX}]$	-0.3285	-0.009429	-0.003648	-0.004302	-0.003251	0.002673	0.5543	-0.003344	0.03196	-0.004756	-0.00171
Panel B: skewness											
$S_{\gamma^J}(1/2)$	-0.06286	-0.0001898	4.38e-06	-3.979e-06	1.304e-05	0.0002412	0.1284	5.945e-06	0.004474	-0.0001434	0.0003139
$\mathcal{C}[S_{\gamma^J}(1/2)]$	-5.67	-0.03245	-0.005487	-0.00708	-0.003578	0.02172	11	0.004335	0.4408	-0.0125	0.02978
$\mathcal{P}[S_{\gamma^J}(1/2)]$	-11.01	-0.02218	0.005413	0.00354	0.007231	0.03271	5.673	-0.004329	0.4419	-0.02954	0.01217
$\mathcal{T}[S_{\gamma^J}(1/2)]$	-12.69	-0.1389	-0.05752	-0.06211	-0.05372	-0.02084	8.186	-0.1527	0.6282	-0.187	-0.1275
Panel C: quarticity											
$Q_{\gamma^J}(1/2)$	1.71e-09	3.616e-06	1.306e-05	1.09e-05	1.564e-05	5.551e-05	0.06128	0.0002529	0.002172	0.0001807	0.0004353
$\mathcal{C}[Q_{\gamma^J}(1/2)]$	-12.05	-0.01531	-0.0004048	-0.001003	0.0001241	0.007236	6.577	-0.0233	0.5288	-0.05544	-0.00193
$\mathcal{P}[Q_{\gamma^J}(1/2)]$	-6.569	-0.007201	0.0004722	-0.0001067	0.00101	0.01531	12.05	0.02356	0.5285	0.001757	0.05589
$\mathcal{T}[Q_{\gamma^J}(1/2)]$	-18.81	-0.1136	-0.03534	-0.04063	-0.0317	-0.0109	8.237	-0.2049	1.009	-0.2702	-0.1642

Descriptive statistics of open-to-close (8:30 to 15:15) 5-minute frequency divergence trading profits with weekly option settlement for strategy γ^J. Trading period: from 2010-12-30 to 2015-12-30 with a total of 1258 days. $\mathcal{C}[\cdot]$ denotes the integrated option cost at mid prices; $\mathcal{P}[\cdot]$ denotes total trading profits at mid prices; $\mathcal{T}[\cdot]$ denotes total trading profits after the transaction costs (bid-ask spread) are taken into account. $\mathcal{P}[\cdot]$ and $\mathcal{T}[\cdot]$ are reported in the following way: if mean $\mathcal{P}[\cdot]$ from a *long* strategy is negative, $\mathcal{P}[\cdot]$ and $\mathcal{T}[\cdot]$ from a *short* strategy are reported. q_x denotes $100x$th percentile; Med denotes median; SD denotes standard deviation; MAD denotes median absolute deviation.

TABLE A.6: Summary statistics of close-to-open divergence trading with weeklies

qty	Min	$q_{0.25}$	Med	CI (median) 90%		$q_{0.75}$	Max	Mean	SD	CI (mean) 90%	
Panel A: divergence											
$D_{\gamma^J}(1/2)$	7.392e-08	0.0003514	0.00155	0.001441	0.001774	0.005503	0.1386	0.005493	0.01169	0.00497	0.006158
$\mathcal{C}[D_{\gamma^J}(1/2)]$	-0.07597	0.002953	0.00555	0.005194	0.005915	0.01093	0.2602	0.009274	0.01782	0.008477	0.01033
$\mathcal{P}[D_{\gamma^J}(1/2)]$	-0.2285	-0.006935	-0.003535	-0.003756	-0.003305	-0.0009259	0.2034	-0.003781	0.01885	-0.004744	-0.002836
$\mathcal{T}[D_{\gamma^J}(1/2)]$	-0.3184	-0.005829	-0.0007378	-0.001117	-0.000436	0.001779	0.166	-0.005119	0.02181	-0.006384	-0.004137
D_{VIX}	8.525e-07	0.002267	0.01095	0.007766	0.01307	0.0272	0.3241	0.02312	0.03849	0.01933	0.02882
$\mathcal{C}[D_{VIX}]$	0.01637	0.02971	0.04494	0.04062	0.04792	0.06418	0.5256	0.06586	0.07513	0.05824	0.07647
$\mathcal{P}[D_{VIX}]$	-0.4315	-0.04626	-0.02772	-0.03227	-0.02542	-0.01633	0.06974	-0.04273	0.06378	-0.05136	-0.03611
$\mathcal{T}[D_{VIX}]$	-0.1043	0.01143	0.0226	0.01969	0.02521	0.0381	0.3883	0.03297	0.05601	0.02705	0.04059
Panel B: skewness											
$S_{\gamma^J}(1/2)$	-0.7087	-0.0002245	2.721e-06	2.293e-07	6.231e-06	0.0005384	0.5688	-0.0001522	0.03928	-0.002119	0.001611
$\mathcal{C}[S_{\gamma^J}(1/2)]$	-4.45	-0.06159	-0.0239	-0.02629	-0.02239	-0.007998	2.09	-0.05744	0.2575	-0.07111	-0.04652
$\mathcal{P}[S_{\gamma^J}(1/2)]$	-2.799	0.00737	0.02474	0.02285	0.02745	0.0643	4.483	0.05728	0.2774	0.04472	0.07122
$\mathcal{T}[S_{\gamma^J}(1/2)]$	-3.262	-0.02571	0.001545	0.0002621	0.002658	0.01861	3.208	-0.01626	0.2439	-0.0278	-0.004963
Panel C: quarticity											
$Q_{\gamma^J}(1/2)$	8.806e-15	1.118e-07	9.929e-07	8.095e-07	1.3e-06	8.611e-06	0.004213	4.876e-05	0.0002752	3.677e-05	6.683e-05
$\mathcal{C}[Q_{\gamma^J}(1/2)]$	-0.2191	0.0006619	0.002072	0.0019	0.002289	0.006414	0.7328	0.008294	0.03882	0.006689	0.01088
$\mathcal{P}[Q_{\gamma^J}(1/2)]$	-0.7325	-0.006362	-0.002072	-0.002278	-0.001896	-0.0006347	0.2221	-0.008245	0.03883	-0.01079	-0.006484
$\mathcal{T}[Q_{\gamma^J}(1/2)]$	-0.3063	-0.0002681	0.0003993	0.0003391	0.0004561	0.00174	0.4742	-0.0001248	0.02795	-0.001597	0.001335

Descriptive statistics of close-to-open (15:15 to 8:30) divergence trading profits with weekly option settlement for strategy γ^J. 99% of payoffs with smallest absolute replication error are retained. Trading period: from 2010-12-31 to 2015-12-24 with a total of 261 days $\mathcal{C}[\cdot]$ denotes the integrated option cost at mid prices; $\mathcal{P}[\cdot]$ denotes total trading profits at mid prices; $\mathcal{T}[\cdot]$ denotes total trading profits after the transaction costs (bid-ask spread) are taken into account. $\mathcal{P}[\cdot]$ and $\mathcal{T}[\cdot]$ are reported in the following way: if mean $\mathcal{P}[\cdot]$ from a *long* strategy is negative, $\mathcal{P}[\cdot]$ and $\mathcal{T}[\cdot]$ from a *short* strategy are reported. q_x denotes $100x$th percentile; Med denotes median; SD denotes standard deviation; MAD denotes median absolute deviation.

A.2 Appendix to Chapter 2

A.2.1 Tables

$1-\alpha$	Annually	
data	0.00302	
model	−0.01017	
95%	−0.007307	0.01371
99%	−0.01025	0.01703

TABLE A.7: Excess dispersion in the Bansal, Kiku, and Yaron, 2016 model and in the data. Model values calculated with the use of their best estimated model, whose parameters are reported in Table II of their paper. Data values calculated from sample ranging from 1946-03-30 to 2012-10-31. Confidence intervals calculated with the use of a time-series bootstrap (basic confidence interval type).

Bibliography

Acciaio, B. et al. (2013). "A Model-Free Version of the Fundamental Theorem of Asset Pricing and the Super-Replication Theorem". In: *Mathematical Finance*, n/a–n/a. ISSN: 1467-9965. DOI: 10.1111/mafi.12060. URL: http://dx.doi.org/10.1111/mafi.12060.

Aït-Sahalia, Yacine (2004). "Disentangling diffusion from jumps". In: *Journal of Financial Economics* 74.3, pp. 487 –528. ISSN: 0304-405X. DOI: http://dx.doi.org/10.1016/j.jfineco.2003.09.005. URL: http://www.sciencedirect.com/science/article/pii/S0304405X04000777.

Aït-Sahalia, Yacine, Mustafa Karaman, and Loriano Mancini (2015). "The Term Structure of Variance Swaps and Risk Premia". In: *Available at SSRN*.

Almeida, Caio and René Garcia (2017). "Economic Implications of Nonlinear Pricing Kernels". In: *Management Science* 0.0, null. DOI: 10.1287/mnsc.2016.2498. eprint: http://dx.doi.org/10.1287/mnsc.2016.2498. URL: http://dx.doi.org/10.1287/mnsc.2016.2498.

Alvarez, Fernando and Urban J Jermann (2005). "Using asset prices to measure the persistence of the marginal utility of wealth". In: *Econometrica* 73.6, pp. 1977–2016.

Amaya, Diego et al. (2011). "Does Realized Skewness and Kurtosis Predict the Cross-Section of Equity Returns?" English. In: *SSRN eLibrary*.

Andersen, T., V. Todorov, and N. Fusari (2015). "The Risk Premia Embedded in Index Options". In: *Journal of Financial Economics* Forthcoming.

Andersen, Torben, Nicola Fusari, and Viktor Todorov (2015a). "Parametric inference and dynamic state recovery from option panels". In: *Econometrica* Forthcoming.

— (2016). "The Pricing of Short-Term market Risk: Evidence from Weekly Options". In: *Journal of Finance*.

Andersen, Torben G., Tim Bollerslev, and Xin Huang (2011). "A reduced form framework for modeling volatility of speculative prices based on realized variation measures". In: *Journal of Econometrics* 160.1. Realized Volatility, pp. 176 –189. ISSN: 0304-4076. DOI: http://dx.doi.org/10.1016/j.jeconom.2010.03.029. URL: http://www.sciencedirect.com/science/article/pii/S0304407610000710.

Andersen, Torben G., Oleg Bondarenko, and Maria T. Gonzalez-Perez (2015). "Exploring Return Dynamics via Corridor Implied Volatility". In: *Review of Financial Studies* 28.10, pp. 2902–2945. DOI: 10.1093/rfs/hhv033. eprint: http://rfs.oxfordjournals.org/content/28/10/2902.full.pdf+html.

URL: http://rfs.oxfordjournals.org/content/28/10/2902.abstract.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov (2015b). "The risk premia embedded in index options". In: *Journal of Financial Economics* 117.3, pp. 558 –584. ISSN: 0304-405X. DOI: http://dx.doi.org/10.1016/j.jfineco.2015.06.005. URL: http://www.sciencedirect.com/science/article/pii/S0304405X15000987.

Andersen, Torben G. et al. (2015). "The fine structure of equity-index option dynamics". In: *Journal of Econometrics* 187.2. Econometric Analysis of Financial Derivatives, pp. 532 –546. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2015.02.037. URL: http://www.sciencedirect.com/science/article/pii/S0304407615000627.

ANG, ANDREW et al. (2006). "The Cross-Section of Volatility and Expected Returns". In: *The Journal of Finance* 61.1, pp. 259–299. ISSN: 1540-6261. DOI: 10.1111/j.1540-6261.2006.00836.x. URL: http://dx.doi.org/10.1111/j.1540-6261.2006.00836.x.

Backus, David, Nina Boyarchenko, and Mikhail Chernov (2016). *Term Structures of Asset Prices and Returns*. Working Paper 22162. National Bureau of Economic Research. DOI: 10.3386/w22162. URL: http://www.nber.org/papers/w22162.

Backus, David, Mikhail Chernov, and Ian Martin (2011). "Disasters Implied by Equity Index Options". In: *The Journal of Finance* 66.6, pp. 1969–2012. ISSN: 1540-6261. DOI: 10.1111/j.1540-6261.2011.01697.x. URL: http://dx.doi.org/10.1111/j.1540-6261.2011.01697.x.

Backus, David, Mikhail Chernov, and Stanley Zin (2014). "Sources of entropy in representative agent models". In: *The Journal of Finance* 69.1, pp. 51–99.

Bajgrowicz, Pierre, Olivier Scaillet, and Adrien Treccani (2015). "Jumps in High-Frequency Data: Spurious Detections, Dynamics, and News". In: *Management Science* 62.8, pp. 2198–2217. ISSN: 0025-1909. DOI: 10.1287/mnsc.2015.2234. URL: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.2015.2234 (visited on 09/13/2016).

Bakshi, Gurdip and Fousseni Chabi-Yo (2012). "Variance bounds on the permanent and transitory components of stochastic discount factors". In: *Journal of Financial Economics* 105.1, pp. 191–208.

— (2014). "New Entropy Restrictions and the Quest for Better Specified Asset Pricing Models". In: *Charles A. Dice Center Working Paper No. 2014-07* 2014-03, p. 007.

Bansal, R. and Y. Yaron (2004). "Risks for the long run: A potential resolution of asset pricing puzzles". In: *Journal of Finance* 59, pp. 1481–1509.

Bansal, Ravi, Dana Kiku, and Amir Yaron (2016). "Risks for the long run: Estimation with time aggregation". In: *Journal of Monetary Economics* 82, pp. 52 –69. ISSN: 0304-3932. DOI: http://dx.doi.org/10.1016/j.jmoneco.2016.07.

003. URL: http://www.sciencedirect.com/science/article/pii/S030439321630040X.

Bansal, Ravi and Bruce N Lehmann (1997). "Growth-optimal portfolio restrictions on asset pricing models". In: *Macroeconomic Dynamics* 1.02, pp. 333–354.

Barndorff-Nielsen, Ole E. and Neil Shephard (2006). "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation". In: *Journal of Financial Econometrics* 4, pp. 1 –30.

Barndorff-Nielsen, Ole E., Neil Shephard, and Matthias Winkel (2006). "Limit theorems for multipower variation in the presence of jumps". In: *Stochastic Processes and their Applications* 116.5, pp. 796 –806. ISSN: 0304-4149. DOI: 10.1016/j.spa.2006.01.007. URL: http://www.sciencedirect.com/science/article/pii/S0304414906000081.

Bates, David (1996). "Jumps and stochastic volatility: Exchange rate processes implicit in Deutschemark options". In: *Review of Financial Studies* 9, pp. 69–108.

— (2000). "Post-'87 Crash Fears in the S&P 500 Futures Option Market". In: *Journal of Econometrics* 94.1, pp. 181–238.

Bates, David S. (2012). "U.S. stock market crash risk, 1926-2010". In: *Journal of Financial Economics* 105.2, pp. 229 –259. ISSN: 0304-405X. DOI: http://dx.doi.org/10.1016/j.jfineco.2012.03.004. URL: http://www.sciencedirect.com/science/article/pii/S0304405X12000414.

Bollerslev, Tim, Sophia Zhengzi Li, and Viktor Todorov (2016). "Roughing up beta: Continuous versus discontinuous betas and the cross section of expected stock returns". In: *Journal of Financial Economics* 120.3, pp. 464 –490. ISSN: 0304-405X. DOI: http://dx.doi.org/10.1016/j.jfineco.2016.02.001. URL: http://www.sciencedirect.com/science/article/pii/S0304405X16300010.

Bollerslev, Tim and Viktor Todorov (2011a). "Estimation of Jump Tails". In: *Econometrica* 79.6, pp. 1727–1783. ISSN: 1468-0262. DOI: 10.3982/ECTA9240. URL: http://dx.doi.org/10.3982/ECTA9240.

— (2011b). "Tails, fears, and risk premia". In: *The Journal of Finance* 66.6, pp. 2165–2211.

Bondarenko, Oleg (2014). "Variance trading and market price of variance risk". In: *Journal of Econometrics* 180.1, pp. 81 –97. ISSN: 0304-4076. DOI: http://dx.doi.org/10.1016/j.jeconom.2014.02.001. URL: http://www.sciencedirect.com/science/article/pii/S0304407614000207.

Bregman, Lev M. (1967). "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming". In: {*USSR*} *Computational Mathematics and Mathematical Physics* 7.3, pp. 200 –217.

Britten-Jones, Mark and Anthony Neuberger (2000). "Option Prices, Implied Price Processes, and Stochastic Volatility". In: *Journal of Finance* 55, pp. 839–866.

Calvet, Laurent E. et al. (2015). "What is beneath the surface? Option pricing with multifrequency latent states". In: *Journal of Econometrics* 187.2. Econometric Analysis of Financial Derivatives, pp. 498 –511. ISSN: 0304-4076. DOI: http://dx.doi.org/10.1016/j.jeconom.2015.02.034. URL: http://www.sciencedirect.com/science/article/pii/S0304407615000597.

Carhart, Mark M. (1997). "On Persistence in Mutual Fund Performance". In: *Journal of Finance* 52(1), pp. 57–82.

Carr, P. and L. Wu (2009a). "Variance Risk Premia". In: *Review of Financial Studies* 22.3, pp. 1311–1341.

Carr, Peter and Roger Lee (2009). "PUT-CALL SYMMETRY: EXTENSIONS AND APPLICATIONS". In: *Mathematical Finance* 19.4, pp. 523–560. ISSN: 1467-9965. DOI: 10.1111/j.1467-9965.2009.00379.x. URL: http://dx.doi.org/10.1111/j.1467-9965.2009.00379.x.

Carr, Peter and Dilip Madan (1998). "Towards a theory of volatility trading". In: *in R. Jarrow (ed) Risk Book on Volatility. New York: Risk, pp 417-27*.

— (2001). "Optimal positioning in derivative securities". In: *Quantitative Finance*.

Carr, Peter and Liuren Wu (2009b). "Variance Risk Premiums". In: *Review of Financial Studies* 22.3, pp. 1311–1341.

CBOE (2000). *The CBOE Volatility Index – VIX*. Tech. rep. CBOE.

Chabi-Yo, Fousseni and Riccardo Colacito (2013). "The Term Structures of Co-Entropy in International Financial Markets". In: *Charles A. Dice Center Working Paper No. 2013-17* 2013-17.

Chang, Bo Young, Peter Christoffersen, and Kris Jacobs (2009). "Market Skewness Risk and the Cross-Section of Stock Returns". English. In: *SSRN eLibrary*.

Chernoff, H. (1952). "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations". In: *The Annals of Mathematical Statistics* 23, pp. 493–507.

Christensen, Kim, Roel C. A. Oomen, and Mark Podolskij (2014). "Fact or friction: Jumps at ultra high frequency". In: *Journal of Financial Economics* 114.3, pp. 576–599. ISSN: 0304-405X. DOI: 10.1016/j.jfineco.2014.07.007. URL: http://www.sciencedirect.com/science/article/pii/S0304405X14001548 (visited on 09/13/2016).

Christoffersen, Peter et al. (2014). "Nonlinear Kalman Filtering in Affine Term Structure Models". In: *Management Science* 60.9, pp. 2248–2268. DOI: 10.1287/mnsc.2013.1870. eprint: https://doi.org/10.1287/mnsc.2013.1870. URL: https://doi.org/10.1287/mnsc.2013.1870.

Dragomirescu, F and C Ivan (1992). "The smallest convex extensions of a convex function". In: *Optimization* 24.3-4, pp. 193–206.

Duffie, D., D. Filipovic, and W. Schachermayer (2003). "Affine Processes and Applications in Finance". In: *Annals of Applied Probability* 13, pp. 984–1053.

Duffie, Darrell, Jun Pan, and Kenneth Singleton (2000). "Transform analysis and asset pricing for affine jump-diffusions". In: *Econometrica* 68 (6), pp. 1343–1376.

Eraker, Bjørn, Michael Johannes, and Nicholas Polson (2003). "The Impact of Jumps in Equity Index Volatility and Returns". In: *Journal of Finance* 58.3, pp. 1269–1300.

Fama, Eugene F. (1965). "The Behavior of Stock-Market Prices". In: *The Journal of Business* 38.1, pp. 34–105. ISSN: 00219398, 15375374. URL: http://www.jstor.org/stable/2350752.

Fama, Eugene F. and Kenneth French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journals of Financial Econometrics* 33, pp. 3–56.

Fulop, A. and J. Li (2015). "Inferring Volatility Dynamics and Variance Risk Premia: An Efficient Bayesian Approach". In: *Working P*.

Goyal, Amit and Alessio Saretto (2009). "Cross-section of option returns and volatility". In: *Journal of Financial Economics* 94.2, pp. 310 –326. ISSN: 0304-405X. DOI: http://dx.doi.org/10.1016/j.jfineco.2009.01.001. URL: http://www.sciencedirect.com/science/article/pii/S0304405X09001251.

Gruber, Peter H., Claudio Tebaldi, and Fabio Trojani (2015). "The Price of the Smile and Variance Risk Premia". In: *Working Paper*.

Hobson, David and Martin Klimmek (2012). "Model-independent hedging strategies for variance swaps". In: *Finance and Stochastics* 16.4, pp. 611–649. ISSN: 1432-1122. DOI: 10.1007/s00780-012-0190-3. URL: http://dx.doi.org/10.1007/s00780-012-0190-3.

Huang, Xin and George Tauchen (2005). "The Relative Contribution of Jumps to Total Price Variance". In: *Journal of Financial Econometrics* 3.4, pp. 456–499. DOI: 10.1093/jjfinec/nbi025. eprint: http://jfec.oxfordjournals.org/content/3/4/456.full.pdf+html. URL: http://jfec.oxfordjournals.org/content/3/4/456.abstract.

Itakura, F. and S. Saito (1968). "Analysis synthesis telephony based on the maximum likelihood method". In: *Proceedings of the 6th International Congress on Acoustics*. Vol. 17. IEEE, pp. C17–C20.

Jacod, J. and P. Protter (2012). *Discretization of Processes*. Vol. 67. Stochastic Modelling and Applied Probability. Heidelberg: Springer-Verlag.

Jammalamadaka, S. R., T.S. Rao, and G. Terdik (2006). "Higher Order Cumulants of Random VVector and Applications to Statistical Inference and Time Series". In: *Sankhyā: The Indian Journal of Statistics (2003-2007)* 68.2, pp. 326–356.

Jiang, George and Yisong S. Tian (2007). "Extracting Model-Free Volatility from Option Prices: An Examination of the Vix Index". English. In: *Journal of Derivatives*.

Jiang, George J. and Yisong S. Tian (2005a). "The Model-Free Implied Volatility and Its Information Content". In: *Review of Financial Studies* 18.4, pp. 1305–1342.

— (2005b). "The Model-Free Implied Volatility and Its Information Content". In: *Review of Financial Studies* 18.4, pp. 1305–1342. DOI: 10.1093/rfs/hhi027. eprint: http://rfs.oxfordjournals.org/content/18/4/1305.full.pdf+html. URL: http://rfs.oxfordjournals.org/content/18/4/1305.abstract.

Khajavi, Ali Noori, Piotr Orłowski, and Fabio Trojani (2016). "Realized Divergence".

Lee, Roger (2010). "Weighted Variance Swap". In: *Encyclopedia of Quantitative Finance*. John Wiley & Sons, Ltd. ISBN: 9780470061602. DOI: 10.1002/9780470061602.eqf07045. URL: http://dx.doi.org/10.1002/9780470061602.eqf07045.

Li, Jia (2011). *Testing for jumps: A delta-hedging perspective*. Tech. rep. Working paper, Princeton University.

Li, Junye (2013). "An unscented Kalman smoother for volatility extraction: Evidence from stock prices and options". In: *Computational Statistics & Data Analysis* 58. The Third Special Issue on Statistical Signal Extraction and Filtering, pp. 15 –26. ISSN: 0167-9473. DOI: http://dx.doi.org/10.1016/j.csda.2011.06.001. URL: http://www.sciencedirect.com/science/article/pii/S0167947311002015.

Li, Junye and Gabriele Zinna (2017). "The Variance Risk Premium: Components, Term Structures, and Stock Return Predictability". URL: http://dx.doi.org/10.1080/07350015.2016.1191502.

Liu, Jun, Francis A. Longstaff, and Jun Pan (2003). "Dynamic Asset Allocation with Event Risk". In: *The Journal of Finance* 58.1, pp. 231–259. ISSN: 1540-6261. DOI: 10.1111/1540-6261.00523. URL: http://dx.doi.org/10.1111/1540-6261.00523.

Liu, Yan (2015). "Index Option Returns and Generalized Entropy Bounds". In: *Available at SSRN 2149265*.

Mancini, Cecilia (2001). "Disentangling the jumps of the diffusion in a geometric jumping brownian motion". In: *Giornale dell'Istituto Italiano degli Attuari* LXIV, pp. 19–47.

Martin, Ian (2012). *Simple Variance Swaps*. Working Paper 16884. National Bureau of Economic Research. URL: http://www.nber.org/papers/w16884.

— (2013). *Simple Variance Swaps*. Working Paper 16884. National Bureau of Economic Research.

Merwe, Rudolph van der and Eric A. Wan (2001). "The square-root unscented Kalman filter for state and parameter-estimation". In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP '01)*. Vol. 6, pp. 3461–3464. DOI: 10.1109/ICASSP.2001.940586.

Moore, A. (1962). "A statistical analysis of common stock prices". PhD thesis. University of Chicago, Graduate School of Business.

Muravyev, Dmitriy and Xuechuan Ni (2016). *Why Do Option Returns Change Sign from Day to Night?* SSRN Scholarly Paper ID 2820264. Rochester, NY: Social Science Research Network.

Neuberger, Anthony (1994). "The Log Contract". In: *Journal of Portfolio Management* 20.2, pp. 74–80.

Orłowski, Piotr (2015). "Big risk". In: Working paper, Usi Lugano and SFI.

Orłowski, Piotr, András Sali, and Fabio Trojani (2015). "Arbitrage Free Dispersion". manuscript, University of Lugano.

Pan, Jun (2002). "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study". In: *Journal of Financial Economics* 63.1, pp. 3–50.

Peters, Hans and Peter Wakker (1987). "Convex functions on non-convex domains". In: *Economics letters* 22.2-3, pp. 251–255.

Press, S. James (1967). "A Compound Events Model for Security Prices". In: *The Journal of Business* 40.3, pp. 317–335. ISSN: 00219398, 15375374. URL: http://www.jstor.org/stable/2351754.

Rényi, A. (1960). "On measures of information and entropy". In: *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability*, pp. 547–561.

Schneider, Paul G. and Fabio Trojani (2015a). "Divergence and The Price of Uncertainty". Working Paper.

— (2015b). "Fear Trading". English. Swiss Finance Institute Research Paper No. 15-03. http://ssrn.com/paper=1994454.

Snow, Karl N (1991). "Diagnosing asset pricing models using the distribution of asset returns". In: *The Journal of Finance* 46.3, pp. 955–983.

Tzoumas, V., A. Jadbabaie, and G. J. Pappas (2016). "Sensor placement for optimal Kalman filtering: Fundamental limits, submodularity, and algorithms". In: *2016 American Control Conference (ACC)*, pp. 191–196. DOI: 10.1109/ACC.2016.7524914.

Wan, Eric A. and Rudolph van der Merwe (2000). "The unscented Kalman filter for nonlinear estimation". In: *Proc. and Control Symp Adaptive Systems for Signal Processing, Communications 2000. AS-SPCC. The IEEE 2000*, pp. 153–158. DOI: 10.1109/ASSPCC.2000.882463.